

„... această carte este prima de genul ei în limba română și este îmbucurător faptul că ideea elaborării ei vine de la un tînăr cercetător statistician, pasionat de meseria sa și în mod special de aplicațiile practice ale științei statistice."

Acad. GH. MIHOC

Lei 5,25



SINTEZE LYCEUM

VIOREL GH. VODĂ

GÎNDIREA STATISTICĂ — UN MOD DE GÎNDIRE AL VIITORULUI

EDITURA ALBATROS 

Coperta colecţiei de
ANDREI OLSUFIEV

Referent ştiinţific:
Ing. DAN ILIESCU

BUCUREŞTI • 1977



SINTEZE
LYCEUM

VIOREL GH. VODĂ

GÎNDIREA STATISTICĂ — UN MOD DE GÎNDIRE AL VIITORULUI

Cuvînt înainte de Acad. GH. MIHOC

EDITURA ALBATROS



CUVÎNT ÎNAINTE

Cărți asupra metodelor specifice diferitelor științe s-au scris multe. O bogată literatură există și în domeniul statisticii. Sînt însă rare cărțile despre gîndirea statistică, despre locul statisticii în contextul celorlalte științe, despre valoarea ei filozofică și fundamentalele ei aplicații în toate ramurile de activitate umană. După cunoștința mea, această carte este prima în genul ei în limba română și este îmbucurător faptul că ideea elaborării ei vine de la un tînăr cercetător statistician, pasionat de meseria sa și în mod special de aplicațiile practice ale științei statistice.

În timpul nostru, statistica a realizat ca știință teoretică progrese spectaculoase, la același nivel cu toate ramurile matematicii, putînd satisface exigențele celui mai pretențios cercetător. Autorul, deși a publicat în statistica teoretică lucrări valoroase, face parte dintre cei care înțeleg că cercetarea, mai cu seamă în statistică, trebuie să servească unui scop precis: rezolvarea unor probleme de mare importanță, indiferent dacă ele sînt de natură economică, industrială, de agricultură sau din viața socială.

Documentele Partidului și Statului nostru au subliniat cu deosebită justețe faptul că știința, cercetarea științifică, proiectarea și învățămîntul, lucrînd în strînsă colaborare, trebuie să aducă în continuare o contribuție și mai substanțială la progresul întregii societăți.

Cincinalul revoluției tehnico-științifice creează condițiile desfășurării unei activități rodnice pentru toți oamenii muncii, în vederea ridicării societății românești pe o nouă treaptă materială și spirituală.

Tînăra generație are ce învăța din această carte: întîi despre statistică, despre nenumăratele ei aplicații și apoi despre activitatea de cercetare dusă în folosul Patriei Socialiste.

Acad. GH. MIHOC

DIN PARTEA AUTORULUI

Ideea de a scrie o astfel de carte este mai veche. Întîmplător — cuvînt preferat al statisticienilor — ea coincide cu anul absolvirii Facultății de matematică-mecanică a Universității București, secția Teoria probabilităților și statistica matematică.

Într-o zi din acel an, întrebat fiind ce meserie am, răspunsesem fără prea multe ezitări: statistician. Reacția partenerului a fost promptă: Aha! Ești din cei care adună pe linii și coloane și le dă la fel!

Deși pretenția de a mă proclama atunci statistician era exagerată — ținînd cont de ceea ce se înțelege azi prin statistician — totuși am replicat (fără însă a putea să dovedesc prea bine acest lucru) că statistica nu înseamnă de fel adunări pe linii și coloane, ci în primul rînd o concepție nouă de gîndire, cea mai adecvată în studierea fenomenelor din diverse domenii de activitate.

Neputînd să demonstrez ceea ce intuitiv simțeam că e adevărat, mi-a apărut atunci un aspect nou: cursurile unei facultăți îți oferă o calificare potențială; ceea ce convertește această calificare potențială într-o reală profesie, într-o adevărată brățară de aur, este practica. De la acea întîmplare a trecut un deceniu.

Legăturile colectivului de statistică industrială de la Centrul de statistică matematică din București, unde am lucrat tot acest răstimp, cu întreprinderi industriale de cea mai diversă specializare, începînd cu Uzinele de fibre chimice din Săvinești — adevărată școală

a tinerilor statisticieni ce eram — ne-au oferit, credem, rîvnita brățară din metalul nobil al practicii.

Vizitele și documentările ulterioare prin diferite țări din lume n-au făcut decît să ne confirme acest deziderat al importanței excepționale a contactului cu realitatea în activitatea cercetătorului din domeniul statisticii matematice.

Astfel, în mod firesc, ideea integrării: învățămînt-cercetare-producție, nu poate decît să vină în ajutorul tinerilor care se dedică statisticii matematice.

Cercetarea științifică în acest domeniu poate și trebuie să ajute și mai mult economia țării noastre în vederea edificării societății socialiste multilateral dezvoltate.

Tovarășul Nicolae Ceaușescu la cel de-al XI-lea Congres al Partidului Comunist Român, a arătat că „... *putem asigura creșterea contribuției științei la înfăptuirea programului de modernizare și ridicare a nivelului tehnic și calitativ al întregii noastre economii, de perfecționare a întregii vieți sociale*"[1].

Partidul și statul acordă o deosebită atenție cercetării științifice, în general. Astfel, în perioada actualului cincinal, în domeniul aplicațiilor matematicii în economie — se va urmări perfecționarea metodelor statistice și de programare matematică pentru îmbunătățirea planificării și conducerii la nivel macro și microeconomic, pentru proiectarea utilajelor complexe; perfecționarea metodelor de control al calității și fiabilității produselor.

Există deci condiții extrem de propice desfășurării unei activități cu adevărat utile societății: aplicarea practică imediată a cunoștințelor științifice.

Cartea aceasta a fost scrisă cu mai multe scopuri: în primul rînd să încerce să ofere tinerilor cititori o imagine pe cît posibil de reală a ceea ce este statistica și a ceea ce se poate face cu ea; în al doilea rînd, să aducă un suflu mai larg pentru cei care deja se ocupă cu statistica matematică, dar care prinși în vîltoarea unui domeniu specific acesteia au uitat de unitatea și caracterul interdisciplinar al cunoașterii în epoca noastră. În fine, în al treilea rînd, să trezească un interes mai viu pentru această disciplină aparent aridă și plictisitoare.

---

[1] Ceaușescu, Nicolae, *Raport la cel de-al XI-lea Congres al Partidului Comunist Român*, Editura politică, București, 1974, p. 60.

INTRODUCERE

„Probabil că şi dumneavoastră aţi început să studiaţi statistica — nu pentru că aţi dorit-o, ci pentru că aţi avut nevoie de ea. Dacă aşa stau lucrurile, ştiu cam ce simţiţi. Am trecut şi eu printr-o experienţă asemănătoare cîţiva ani în urmă; dacă ar fi trebuit să evit statistica pesemne că aş fi făcut-o.

Totuşi, atitudinea mea s-a schimbat după ce am început s-o studiez fiindcă am descoperit în ea UN NOU MOD DE GÎNDIRE, care m-a fascinat puternic".

John L. Phillips Jr.[1]

Se pare că scriitorul englez H. G. Wells a fost printre primii nespecialişti în domeniul statisticii care a sesizat, încă acum cîteva decenii, evoluţia rapidă pe care o va avea această ramură a matematicii. Într-o frază rămasă celebră, şi citată azi cu mîndrie de statisticieni, el spunea că în viitorul nu prea îndepărtat oamenii vor trebui să înveţe statistică aşa cum învaţă să scrie şi să citească.

Multă vreme, remarca lui Wells a părut exagerată, însă evoluţia ulterioară a ştiinţei şi tehnicii, a societăţii în general, a dovedit că el nu era deloc departe de adevăr.

Cum se explică totuşi că numai în ultimele trei decenii statistica matematică a devenit atît de necesară şi se aplică în atît de multe domenii, deşi teoria probabilităţilor (despre care se poate spune că este baza matematică a statisticii) există de cîteva sute de ani?

Răspunsul la întrebare nu este chiar evident şi el trebuie căutat în evoluţia istorică atît a matematicii

[1] Phillips Jr., J. L., *Statistical Thinking*, W. N. Freeman Co., San Francisco, 1973, p. 1.

în general, în specificul statisticii matematice, cît și în schimbările rapide care au avut loc în dezvoltarea societății umane în secolul nostru. Să încercăm să scoatem în evidență cîteva trăsături specifice ale acestei evoluții.

Este binecunoscut faptul că în trecutul îndepărtat, problemele ridicate de practică au condus treptat la apariția matematicii — știința care astăzi a căpătat un caracter de universalitate.

Clasicii materialismului dialectic au analizat în profunzime unele aspecte legate de aceste probleme. Astfel, Fr. Engels (*Anti-Dühring*, Ed. a III-a, ESPLP, București, 1955, p. 48) sublinia faptul că la fel ca și „... toate celelalte științe, matematica s-a născut din necesitățile practice ale oamenilor: din măsurarea loturilor de pămînt și a capacității vaselor, din calcularea timpului și din mecanică".

Într-adevăr, calendarele aztece, piramidele din Egipt, construcțiile babilonienilor au necesitat calcule matematice ce pot fi considerate ca primele episoade ale începuturilor matematicii ca știință.

În acele timpuri, teoria (redusă la principii de calcul, la elemente de geometrie) era generată aproape în exclusivitate de necesitatea rezolvării unor cerințe practice. Este o etapă în dezvoltarea matematicii, pe care o putem numi *primară*, în care conceptele matematice, teoria în general era efectul direct al problemelor concrete.

Urmează apoi o etapă ce acoperă în timp multe secole și pe care o putem numi *o etapă de dezvoltare „per se"* (pentru sine) în care s-a trecut de la necesități imediate, la imaginarea unor situații mai complexe decît cele reale, la construirea unor concepte și teorii fără aparentă aplicabilitate (unele chiar fără aplicabilitate!), teorii și concepte folosite pentru rezolvarea unor probleme cu caracter pur teoretic.

Este lunga perioadă de avînt a așa-numitei matematici pure, a cărei viteză de dezvoltare a depășit cu mult posibilitățile de aplicare practică a rezultatelor obținute.

Această perioadă a conținut exagerările ei în care s-a ajuns la situația paradoxală de a se crea teorii și apoi

a căuta domeniile practice în care acestea să-și găsească aplicabilitate.

Tot Engels a fost cel care a sesizat pericolul la care este supusă o teorie ce se dezvoltă ignorînd lumea reală. „...ca și în toate domeniile gîndirii, pe o anumită treaptă de dezvoltare, legile abstrase din lumea reală sînt despărțite de lumea reală și îi sînt opuse ca ceva de sine stătător, ca niște legi venind din afară" (idem, op. cit., p. 48).

Dezvoltarea societății a impus matematicii *o etapă de revenire* care poate fi considerată ca fiind epoca modernă, în care evoluția multilaterală a societății și în special sfera activității economice a reclamat găsirea unor metode și căi din ce în ce mai complexe pentru rezolvarea problemelor practice de care depinde însuși progresul societății.

S-a trecut astfel, în mod natural, la reevaluarea întregii cercetări științifice, a rezultatelor sale, efectuîndu-se o adevărată epurare a faptelor minore și a teoriei de dragul teoriei. Așa se poate explica, pe de o parte, și reapariția unor teorii ce păreau uitate (de exemplu, algebra booleană cu largi aplicații în teoria mașinilor de calcul) și, pe de altă parte, trecerea în anonimat a unor ramuri ce azi nu mai prezintă decît un interes pur istoric.

Statistica matematică și-a făcut apariția într-o perioadă istorică în care societatea era deja aptă pentru a-i utiliza rezultatele, astfel că spre deosebire de algebra booleană, de exemplu, ea nu a mai fost obligată să rămînă multă vreme izolată în sfera teoriei.

De asemenea, caracterul extraordinar de complex al problemelor legate de conducerea societății moderne a impus, în mod necesar, acel instrument de studiu specific *fenomenelor de masă*, indiferent că este vorba de industrie, agricultură, viața socială etc.

Ce s-a întîmplat însă cu teoria probabilităților în acest timp? Deși legile acestei teorii au fost puse încă în secolul al XVII-lea — prima carte de teoria probabilităților fiind scrisă (în limba latină) de către fizicianul Huygens, în 1657, tratînd elemente probabilistice ale unor jocuri de noroc, deși Pascal și Fermat, în

1654, şi chiar mai de timpuriu, matematicieni italieni ai secolului al XVI-lea au atacat probleme similare, punîndu-şi, de asemenea, şi chestiuni de metodologie —, se poate considera că abia cu enormul tratat scris de Laplace în 1812 (*Théorie Analytique des Probabilités* — Teoria analitică a probabilităţilor), ce însumează lucrările predecesorilor precum şi contribuţiile sale proprii, teoria probabilităţilor se poate considera în expansiune şi poate marca totodată şi începuturile statisticii matematice. În acest tratat, Laplace a înglobat şi o lucrare a sa — azi celebră — numită *Un eseu filozofic asupra teoriei probabilităţilor* din care putem să ne dăm seama care era concepţia sa, cît şi a matematicienilor epocii asupra acestei ramuri a matematicii.

Întreaga lucrare a lui Laplace este pătrunsă de ideea aplicării matematicii (în special a ceea ce numim noi astăzi *Analiza matematică*) la teoria probabilităţilor.

În *eseu*, Laplace atacă pentru prima oară problema statisticii matematice, privită şi din punct de vedere practic. Laplace subliniază caracterul dualist al dezvoltării statisticii: o parte vizînd considerente practice, economice, iar cealaltă parte vizînd validitatea ei în deducţiile ştiinţifice.

Evoluţia ulterioară a confirmat, după părerea noastră, concepţia lui Laplace. Într-adevăr, de pildă, repartiţia normală — studiată azi şi în ultimul an de liceu (numită şi repartiţia Gauss-Laplace) —, a fost folosită pentru modelarea unor situaţii practice, prima aplicaţie găsindu-şi-o în teoria erorilor, cu evidente aplicaţii şi importanţă practică (de aici şi denumirea folosită cîteodată de „repartiţia erorilor").

La rîndul ei, repartiţia erorilor a dus la concluzia că ea este adecvată şi ca model în teoria cinetică a gazelor — model al repartiţiei vitezei moleculelor.

Secolul al XIX-lea înregistrează importante cuceriri în domeniul statisticii matematice, cuceriri impulsionate în parte şi de dezvoltarea generală a ştiinţelor şi în special a matematicii.

Un fapt trebuie cu deosebire subliniat: instrumentul matematic este indispensabil pentru furnizarea unor

baze logice, riguroase teoriei statisticii. Dar m a t e m a t i c a p u r ă n u a c o n d u s l a c o n c e p t e n o i î n ş t i i n ţ a s t a t i s t i c ă. Acestea nu au fost obţinute decît prin observarea şi investigarea lumii reale, matematica furnizînd tehnicile de calcul.

Exemple în sprijinul acestei idei sînt nenumărate: să menţionăm doar ideea fizică de corelaţie între două varietăţi biologice, care a fost formulată pe baza măsurătorilor asupra unor varietăţi de mazăre dulce, precum şi corelaţia între înălţimea unui grup de indivizi şi cea a descendenţilor lor, observată de Galton la sfîrşitul secolului al XIX-lea.

Treptat, aceste noţiuni şi-au dovedit o aplicabilitate generală, azi ideea de corelaţie statistică fiind des folosită în studiul caracteristicilor proceselor industriale.

Un pas important în evoluţia statisticii a fost apariţia în Anglia a revistei „Biometrika" (la sfîrşitul secolului al XIX-lea) care a încurajat aplicarea metodelor statistice în biologie, medicină, genetică etc. Şcoala engleză de statistică a avut un rol de pionerat în acest domeniu. Mari statisticieni precum R. A. Fisher (considerat azi drept titanul statisticii), K. Pearson, Fr. Yates şi alţii au îmbinat în activitatea lor munca de cercetare cu activitatea practică, fapt care arată încă o dată că ideile mari, cu adevărat importante nu se pot făuri decît în contact cu problemele reale, concrete, ale lumii înconjurătoare.

Astfel, la începutul secolului al XX-lea, statisticianul englez W. Gosset (cunoscut sub pseudonimul „Student") a lucrat într-o fabrică de bere; acolo, punîndu-se problema comparării calităţii diferitelor tipuri de bere, el a ajuns la formularea cunoscutei sale metode (testul *Student*) prin care se evaluează statistic omogenitatea calităţii medii a două sortimente de produse. Un alt statistician englez, R. A. Fisher, şi-a desfăşurat activitatea, timp de mai mulţi ani, la Staţiunea agricolă experimentală de la Rothamsted unde, în 1919, prin studierea unor experimente în agricultură a ajuns la punerea bazelor ştiinţifice ale unei noi discipline statistice numită *Planificarea experimentelor* care azi are

aplicații deosebit de utile în toate domeniile de activitate.

Exemple de acest gen sînt multe. Statistica matematică constituie un exemplu pregnant că cercetarea așa-numită „fundamentală" și cea așa-numită „aplicativă" sînt indivizibile.

Cercetarea urmărește descoperirea de fapte noi, iar problemele puse de însăși dezvoltarea științei și tehnicii constituie un puternic factor stimulator al acestei activități. Evoluția istorică a statisticii precum și tendințele actuale confirmă justețea acestor idei.

Ne-am întrebat, mai înainte, ce s-a întîmplat cu teoria probabilităților în primele decade ale secolului al XX-lea. Aceasta a continuat să se dezvolte în trei direcții principale:

— fundamentele logice ale teoriei probabilităților;
— teoria matematică a probabilităților;
— aplicații.

În primele două direcții, școala sovietică și-a adus o contribuție importantă, prima prezentare sistematică a teoriei probabilităților pe o bază axiomatică fiind făcută în 1933 de către A. N. Kolmogorov într-o lucrare — clasică azi — intitulată *Fundamentele teoriei probabilităților*.

Lucrarea lui Kolmogorov a generat o efervescență creatoare în acest domeniu, în special în ceea ce privește formalizarea matematică.

Faptul că instrumente matematice ca teoria mulțimilor, algebra, analiza funcțională și-au găsit un cîmp de aplicare în teoria probabilităților a făcut pe unii matematicieni să considere teoria probabilităților o simplă aplicație a unor teorii matematice mai elevate (ca, de exemplu, teoria măsurii).

Această părere este simplistă și, într-un fel, extremistă: evenimentele — noțiune de bază în teoria probabilităților — nu sînt simple mulțimi. În spatele formalizării matematice se ascunde fenomenul real, modelat, și neglijarea acestui aspect duce la o investigație sterilă.

Între timp, statistica matematică impulsionată de dezvoltarea industrială, de introducerea mecanizării și

automatizării pe scară tot mai largă, a început să evolueze independent de teoria probabilităților (dar fără a o nega) tocmai sub presiunea rezolvării unor probleme practice deosebit de importante. Astfel, în primele decade ale secolului nostru, problema conducerii și dirijării proceselor tehnologice de mare serie a ridicat, printre altele, și aspectul deosebit de acut al controlului calității produselor.

Școala americană de statistică își aduce o contribuție majoră în acest domeniu: ia naștere *controlul statistic al calității produselor* — singura metodologie de control viabilă în condițiile producției de mare serie.

În a patra decadă a secolului nostru, inginerul american W. Shewhart lansează concepția sa revoluționară asupra investigării și conducerii proceselor tehnologice, concepție care se bazează pe știința statisticii. Ceea ce a surprins pe mulți a fost faptul că metodologia shewhartiană se bazează, în esență, pe un instrument statistico-matematic destul de simplu.

Iarăși reiese pregnant în evidență faptul că statistica matematică doar în îmbinare cu practica poate furniza rezultate cu adevărat valoroase: gîndirea tehnică joacă, în final, un rol hotărîtor.

După cel de-al II-lea război mondial, statistica ia o amploare deosebită tocmai datorită complexității problemelor lumii moderne. Metodele statistice sînt acum intensiv folosite în noi domenii de cercetare științifică precum și în conducerea proceselor economice. Metodologiile de control statistic al calității produselor, de studiere și îmbunătățire a siguranței în funcționare a acestor produse se răspîndesc aproape în întreaga lume tocmai datorită dezvoltării industriale la scară mondială.

„Invazia" de statistică nu este deci întîmplătoare în ultimele decenii: problemele lumii de azi sînt diferite de cele ale lumii de acum 50 de ani. Au apărut probleme cu totul noi, specifice, ca poluarea, secarea rezervelor unor materii prime etc., încît conducerea optimă a societății se pune cu acuitate.

După cum vom vedea, metodele statistico-matematice fac parte integrantă din metodologia de conducere a economiei, de a cărei structură, grad de dezvoltare

materială și socială depinde, în esență, bunăstarea oricărui popor.

În țara noastră, metodele înaintate, viabile în orice domeniu, s-au bucurat întotdeauna de prețuire: succesul politicii Partidului Comunist Român a constat întotdeauna în adaptarea la condițiile concrete ale țării noastre a celor mai înaintate cuceriri ale gîndirii, ale științei și tehnicii mondiale. Este unul din exemplele cele mai pregnante de aplicare creatoare a concepției materialismului dialectic și istoric la specificul țării noastre.

Este interesant, în acest context, să urmărim pe scurt evoluția istorică a statisticii la noi în țară precum și în alte părți.

Subiectul este, prin natura lui, deosebit de vast. Numai istoria dezvoltării statisticii într-o singură țară poate face subiectul mai multor volume. În epoca actuală, fiecare țară are organisme specializate de statistică, comisii naționale de demografie, reviste de statistică, institute de cercetare etc., astfel încît este efectiv imposibil să se cuprindă în cîteva pagini o activitate atît de bogată la scară mondială.

În anul 1969, în Editura DCS (Direcția Centrală de Statistică) a apărut o culegere de articole intitulată *Din istoria statisticii românești*, unde este cuprinsă aproape toată evoluția și dezvoltarea statisticii, sub multiplele ei aspecte, în țara noastră. „Centum decem anni Statisticae Romaniae“ se numește sugestiv articolul semnat în această culegere de către profesorul Manea Mănescu.

Începuturile statisticii în România datează deci de mai bine de un secol. Pionierii statisticii românești au avut însă de înfruntat greutăți, azi de neimaginat, printre care mai importante le-au constituit: lipsa unui învățămînt statistic organizat, lipsa de documentație și, bineînțeles, lipsa de cadre calificate care să poată răspîndi și aplica metodele statistice în diferite domenii. Este important să menționăm că, efectiv, statistica matematică s-a predat organizat în țara noastră, în instituțiile de învățămînt superior, abia după Eliberare.

Înainte de 23 August 1944, din inițiativa unui grup de entuziaști și pasionați statisticieni s-a format Școala de statistică (în anul 1931; ulterior Institutul de statistică) care a reușit să reunească un nucleu de valoroși matematicieni ai timpului, care se ocupau cu teoria probabilităților și statistica matematică: Octav Onicescu, Gheorghe Mihoc, Șerban Gheorghiu, N. Georgescu-Roegen (econometrician de origine română, azi de faimă mondială, profesor emerit la Universitatea Vanderbilt, Nashville, S.U.A.), Nicolae Praporgescu (care a fost printre primii români ce au scris articole de statistică matematică) etc. Această Școală de statistică încerca să suplinească ceea ce învățămîntul din țara noastră în acele timpuri nu reușea să facă: propagarea metodelor statistice și inițierea unei activități de cercetare în statistică.

S-au luat unele inițiative de publicare a unor cărți în domeniul statisticii și al aplicațiilor sale. Astfel, în 1943 apare *Tratatul de matematici actuariale* al profesorului Gh. Mihoc, una dintre primele cărți de matematici aplicate din țara noastră. S-a alcătuit Biblioteca Institutului, cu scopul de a veni în ajutorul celor interesați să se documenteze și s-au ținut cursuri de statistică matematică, demografie, statistică economică etc.

Lipsit de fonduri și mizînd pe bunăvoința (care se producea după legea evenimentelor rare) unor foruri de stat sau particulare, Institutul de statistică n-a reușit să-și dezvolte acțiunile pe plan național. Activitatea meritorie a Institutului n-a putut depăși granițele orașului București și, chiar aici, datorită neînțelegerii manifestate de oficialități, au existat multe greutăți și neajunsuri. Nu s-a putut astfel înființa o publicație periodică destinată în exclusivitate statisticii, așa cum se proceda încă de multă vreme în destule țări din lume. Exista, de asemenea, un dezinteres general față de aplicațiile statisticii în industrie îndeosebi, deoarece țara importa pe atunci majoritatea produselor industriale de bază.

O concluzie care se poate trage din toată această perioadă dinaintea Eliberării este că aplicațiile statis-

ticii în domeniul producției materiale au fost inexistente, în timp ce în alte colțuri de lume, acestea existau încă din primii ani ai secolului nostru.

Iată deci ce handicap uriaș avea de recuperat țara noastră — numai într-o direcție.

În etapa actuală, există o activitate susținută în domeniul statisticii. Aplicațiile statisticii, nu numai că se bucură de atenție deosebită, dar în unele aspecte esențiale — cum ar fi cel al calității produselor — ele sînt o problemă de stat, deoarece astăzi calitatea însăși este o problemă de stat. În legătură cu aceasta, este bine să reamintim că un rol important în răspîndirea metodelor statistice în producție l-a avut și îl are Inspectoratul general de stat pentru controlul calității produselor (I.G.S.C.C.P.).

În colaborare cu alte organisme ca Academia de studii economice, Centrul de statistică matematică, Direcția centrală de statistică, I.G.S.C.C.P. a desfășurat o muncă foarte utilă în ceea ce privește popularizarea și adoptarea celor mai noi metodologii în controlul calității produselor.

La Uzinele de fibre chimice Săvinești au avut loc, în 1971, lucrările primului simpozion pe țară în domeniul statisticii industriale și controlului statistic al calității. Astfel de congrese au mai avut loc apoi la Timișoara și Craiova, iar în cadrul prestigiosului Congres internațional de teoria probabilităților și aplicații, care se ține periodic la Brașov, începînd cu anul 1955, o secție este dedicată aplicațiilor statisticii în industrie.

Din anul 1964 ființează Centrul de statistică matematică, unitate de cercetare gîndită interdisciplinar, care reunește matematicieni și ingineri angrenați în cercetare și aplicații în domenii ca: teoria probabilităților statistică și matematică, controlul calității și fiabilității etc.

În cadrul Academiei de studii economice există o Facultate de statistică industrială, iar astăzi, statistica matematică este predată nu numai în București, ci și în celelalte centre de învățămînt superior din țară: Iași, Cluj-Napoca, Timișoara, Craiova etc.

Biblioteca Academiei R. S. România conține un număr însemnat de lucrări moderne în domeniul statisticii, de cea mai bună calitate.

De-a lungul anilor, statul a acordat burse pentru specializare în domeniul statisticii în țări cu tradiție recunoscută în această direcție, eforturi care au drept scop mărirea contribuției originale românești în această știință — pe de o parte, iar pe de altă parte angrenarea și aplicarea tot mai profundă a acestor cercetări pentru rezolvarea problemelor majore ale economiei naționale.

Toate acestea au la bază faptul că „Partidul pornește în mod constant de la considerentul că știința constituie factorul primordial al progresului contemporan, că societatea socialistă multilateral dezvoltată și comunismul nu pot fi edificate decît pe baza celor mai înaintate cuceriri ale științei și tehnicii" (*Programul P.C.R. de făurire a societății socialiste multilateral dezvoltate și înaintare a României spre comunism*, Editura politică, București, 1975, p. 95).

Activitatea de cercetare în domeniul statisticii și a aplicațiilor sale s-a dezvoltat în fiecare țară din lume, mai rapid sau mai lent, în funcție de diferiți factori specifici care țin în esență de istoria țării respective.

Parcurgînd axa timpului în sens invers, găsim rădăcini ale gîndirii statistice încă din perioada renascentistă. Astfel, matematicianul italian Luca Pacioli (1445—1514) încerca încă în 1494 să rezolve următoarea problemă: care este repartiția echitabilă a sumei între n jucători, sumă pe care trebuie s-o primească unul dintre ei după ce a întrunit un anumit număr de puncte, în cazul în care jocul se întrerupe înainte ca unul dintre parteneri să acumuleze acest punctaj. Fără să știe, Pacioli încerca să rezolve o complicată problemă pe care azi o numim, în termeni moderni, de analiză combinatorie, ramură a matematicii care a generat în bună măsură aparatul matematic al teoriei probabilităților.

Mai tîrziu, Gerolamo Cardano (1501—1576) și Niccolò Tartaglia (1499—1557) — cunoscuți de noi ca acei care au rezolvat ecuația de gradul III „prin radicali", au atacat și ei aceeași problemă, dar fără a-i

găsi rezolvarea. Totuşi, ei au avut meritul de a atrage atenţia asupra acestui gen de probleme, subliniind şi necesitatea unui nou instrument şi a unei noi gîndiri matematice pentru rezolvarea lor.

Bazele acestui instrument — care nu este altceva decît teoria probabilităţilor — sînt puse la peste un secol de la încercările Cardano-Tartaglia prin cercetările unor mari savanţi ai epocii respective, ca Pascal (1623—1662), Fermat (1601—1665), Huygens (1629—1695) şi Jacques Bernoulli (1654—1705), acesta putînd fi considerat fără exagerare drept primul probabilist din lume, în sensul că el a fost primul matematician care s-a ocupat în mod deosebit şi constant de această ramură şi de aplicaţiile ei care abia apăreau (însuşi Bernoulli, în ultimele zile ale vieţii sale studia probleme de demografie prin prisma teoriei probabilităţilor).

Teoria probabilităţilor, cu toate că a acumulat în timp rezultate valoroase, nu şi-a găsit locul cu adevărat ca o teorie unitară şi independentă, decît în perioada modernă (deceniul al patrulea al secolului nostru), odată cu lucrările şcolii sovietice de matematică. În acest sens, este interesant de menţionat că preocupările privind teoria probabilităţilor merg mult în urmă pînă la celebrul geometru rus N. I. Lobacevski (1792 — 1856) care a investigat posibilităţile noii teorii de a ajuta rezolvarea unor probleme complexe de geometrie.

Un rol de seamă în dezvoltarea teoriei probabilităţilor şi statisticii pe plan mondial l-a avut P. L. Cebîşev (1821—1894) — tipul matematicianului enciclopedic, care a lăsat urme de neşters în orice domeniu al matematicii în care a activat. (Vezi, de exemplu, „Gazeta Matematică şi Fizică", Seria B, nr. 3, 1961, pp. 158—159. Este interesant că Cebîşev este autorul a peste 50 de mecanisme ingenioase, printre care şi aritmometrul automat).

Cele două legi esenţiale ale teoriei probabilităţilor şi anume: legea numerelor mari şi teorema limită centrală (vezi, de exemplu, I o s i f e s c u, M. ş.a., *Teoria probabilităţilor şi statistica matematică*, Editura tehnică, Bucureşti, 1966 şi M i h o c, G h., U r s e a n u, V., *Legea numerelor mari — element component*

al legităţii statistice. În: „Revista de Filozofie", tom 12, nr. 6, pp. 20—32), au fost tratate de Cebîşev într-o formă modernă, reuşind să demonstreze extrem de simplu prima lege în condiţii foarte largi. Cea de-a doua lege a fost complet studiată de către doi dintre elevii lui Cebîşev, şi anume A. A. Markov-tatăl (A. A. Markov-fiul este un cunoscut specialist sovietic în domeniul teoriei algoritmilor) şi A. M. Leapunov (1857—1918).

Markov este creatorul unei teorii foarte fecunde numită azi „teoria lanţurilor Markov", ale cărei aplicaţii, începînd cu genetica şi terminînd cu controlul statistic al calităţii produselor, sînt nenumărate. De exemplu, în anul 1967, J. I. Weindling, cunoscut specialist american în domeniul controlului statistic, şi-a susţinut la Universitatea Columbia din New York teza de doctorat cu titlul „Proprietăţile statistice ale unei clase generale de fişe de control a calităţii, tratate ca un proces Markov".

Lucrările academicienilor A. N. Kolmogorov şi N. V. Smirnov, vin să încadreze locul teoriei probabilităţilor în ansamblul ştiinţelor matematice. Cercetările savanţilor sovietici sînt completate de rezultatele oamenilor de ştiinţă italieni Francesco Paolo Cantelli, Guido Castelnuovo, Claudio E. Bonferroni şi Bruno de Finetti. Aproximativ în aceeaşi perioadă (deceniul al patrulea), cunoscutul matematician Vito Volterra (1860—1940) — primul care a dat o înaltă apreciere lucrărilor marelui nostru matematician Traian Lalescu (1882—1929) în domeniul ecuaţiilor integrale (vezi Sacter, O. *Traian Lalescu*. În: Lalescu, T., *Geometria triunghiului*, Editura tineretului, Bucureşti, 1958, pp. 5—9) — atacă sistematic aspectele matematice pe care le impunea tratarea unor probleme de biologie, îndeosebi cele legate de interacţiunea dintre speciile de animale şi plante. Astfel, el publică la Paris în 1931 lucrarea *Sur la Théorie Mathematique de la lutte pour la vie* (Gauthier Villars, Paris, 1931), iar în acelaşi an, în revista italiană „Giornale dell'Istituto Italiano degli Attuari", articolul *Ricerche mathematiche sulle associazioni biologiche* (cercetări matematice asupra grupărilor de vieţuitoare — vol. 2, nr. 3, 1931, pp. 295—355) care pot fi considerate

ca primele lucrări în care se pun bazele abordării statistico-matematice ale problemelor de biometrie.

În paralel cu aceste cercetări teoretice care fac parte din fundamentul probabilistic al metodelor statistice, dezvoltarea industrială a impus cu necesitate pătrunderea rezultatelor teoretice în practică.

Problema „cheie" care începuse să se pună încă de la începutul secolului nostru, era aceea a controlului calității produselor. Metodele probabilistico-statistice erau singurele chemate să rezolve această problemă, deoarece practica dovedise că așa-numitul control 100% — pe lîngă faptul că devenise foarte costisitor, în multe situații el nu se putea aplica datorită naturii distructive a controlului (gîndiți-vă, de exemplu, la controlul conservelor sau al țesăturilor: un control 100% ar însemna distrugerea întregii producții!). Pe de altă parte, s-a mai dovedit că acest tip de control nu furnizează protecția dorită de beneficiar, că prin „filtrul infailibil" al controlului bucată cu bucată scapă multe produse defecte. (Un lucru interesant pentru cititorii noștri: abia controlul 400% elimină toate produsele cu defecte!)

Istoria controlului statistic al calității începe teoretic în anul 1846 cînd cunoscutul matematician rus M. V. Ostrogradski (1801—1861) a publicat un articol intitulat *În problema probabilităților* care, din păcate, nu a fost apreciat la justa lui valoare în epoca respectivă. Se poate considera fără exagerare că *articolul lui Ostrogradski reprezintă prima încercare de a aplica teoria probabilităților în controlul calității produselor.* (vezi Șarapov, I. P., *Primenenie matematiceskoi statistiki v gheologhii — Statisticeskii analiz gheologhiceschih dannih,* Izdatelstvo Nedra, Moskva, 1965, p. 15). Ostrogradskii afirmă că dacă s-ar aplica formulele deduse de el, atunci s-ar putea ușura considerabil munca de verificare a calității pentru un număr foarte mare de baloturi de textile sau saci cu produse cerealiere. În limbaj modern, Ostrogradski își punea dificila problemă a controlului statistic al calității materialelor în vrac.

Primele aplicații efective ale metodelor statistice în controlul calității s-au desfășurat în S.U.A. la Bell Telephone Laboratories (este vorba de E. C. Molina, care în anul 1900 face primele încercări în acest sens). În jurul anilor 1920, H. F. Dodge și H. G. Romig — tot de la BTL, au abordat sistematic problema controlului statistic de recepție al produselor, și au pus la punct o serie de procedee care se aplică și în zilele noastre. Totuși, după cum remarcă un cunoscut specialist în controlul calității, profesorul E. L. Grant, aplicarea metodelor statistice în industrie nu a avut în S.U.A. succesul scontat, de la început. Cei care și-au dat seama de importanța procedeelor statistice în practica industrială, au fost cîțiva experți militari de la Centrul de Cercetări Balistice al Poligonului de la Aberdeen, Statul Maryland, în frunte cu generalul Leslie L. Simon — autorul de mai tîrziu al unei apreciate lucrări de statistică intitulată *Un manual de metode statistice pentru ingineri (An Engineer's Manual Of Statistical Methods,* John Wiley, New York, 1941).

În ceea ce privește metodologia shewhartiană de care am amintit mai înainte (vezi și cap. II) a avut și ea, la început, de întîmpinat o serioasă rezistență. Matematicianul S. S. Wilks (1907—1964) care este considerat azi unul dintre clasicii statisticii, spunea:

„Pînă la începutul războiului (cel de-al doilea război mondial—n.n.) numai cîteva întreprinderi, în special cele ce produceau produse electrice, erau interesate în folosirea procedeelor controlului statistic al calității ... Cu această ocazie, unii producători au descoperit că procente însemnate din produsele lor nu corespundeau cerințelor ... Ei au dorit atunci să îmbunătățească calitatea produselor pe care le fabricau. Mulți s-au îndreptat atunci către procedeele controlului statistic al calității, pe care le-au găsit imediat efective în stabilirea și menținerea controlului proceselor tehnologice" (vezi Wilks, S. S., *Statistical trainingindustry*. În: „Anal. Chem"., vol. 19, 1947, p. 953).

Este interesant să amintim că în anul 1935 apare la Praga cartea profesorului Jaroslav Janko intitulată *Fundamentele inferenței statistice (Zaklad ý Statističke Indukcé,* Praha, 1935) în care era inclus un capitol consacrat aplicațiilor statisticii în controlul calității, expunîndu-se pe larg ideile lui Shewhart la numai cîțiva ani de la publicarea lucrării sale de bază (1931) în domeniul respectiv.

În deceniul al patrulea, în S.U.A., cu colaborarea unui numeros grup de statisticieni din diferite țări ale lumii, au fost elaborate standarde de control statistic de recepție a produselor industriale, care ulterior, cu anumite îmbunătățiri au devenit standarde internaționale. Aceste standarde au fost adoptate oficial și de România în anul 1972 (vezi STAS 3160/72 și STAS 8820/72), punîndu-se astfel bazele unei și mai largi răspîndiri a metodelor statistice în țara noastră.

După cel de-al doilea război mondial, metodele statistice pătrund tot mai profund în practica industrială. În 1946 se înființează ASQC (Asociația Americană pentru Controlul Calității), iar cîțiva ani mai tîrziu, Organizația Europeană pentru Controlul Calității (EOQC) în care România este reprezentată prin Centrul de Statistică Matematică[1].

Succesul metodelor de control statistic al calității este sugestiv ilustrat de industria japoneză, care la sfîrșitul războiului era practic distrusă, iar produsele japoneze antebelice aveau faima tristă de „marfă ieftină și proastă". În 1948 ia ființă Nippon Kagaku Gijutsu Remmei (cunoscută peste hotare sub numele de JUSE — Asociația Oamenilor de Știință și a Inginerilor Japonezi) care începe o acțiune sistematică de mare amploare privind răspîndirea cunoștințelor statistice în rîndul inginerilor, maiștrilor și muncitorilor japonezi, prin intermediul unor cursuri — atît directe cît și la radio și televiziune.

În anul 1950, JUSE ia inițiativa invitării unor mari personalități în domeniul controlului statistic din lumea întreagă, printre care și profesorul american W. E. Deming. Activitatea lui Deming în Japonia a constituit un important catalizator în răspîndirea metodelor și gîndirii shewhartiene în industria japoneză. Ca semn de omagiu pentru această activitate, JUSE a instituit „Premiul Deming" — acordat persoanelor care au contribuții teoretice meritorii în domeniul statisticii aplicate în industrie, precum și acelor întreprinderi care

1 Actualmente, membru în EOQC, ca reprezentant al S. R. România, este I. G. S. C. C. P, în cadrul căruia C. S. M. figurează ca unitate de cercetare.

au obținut succese în domeniul aplicării controlului statistic al calității. În țara noastră, revista „Calitatea Producției și Metrologie" a publicat deseori informații privind activitatea de aplicare a metodelor statistice în industria japoneză, traduceri ale unor lucrări ale lui W. E. Deming, materiale privind diverși laureați ai premiului Deming, subliniind rolul de „marcă de calitate" pe care îl joacă acest premiu în industria japoneză (vezi, de exemplu, vol. II (XIX), nr. 5, 1972, pp. 288—296 al revistei mai sus amintite, unde se găsește textul integral al lucrării lui W. E. Deming intitulată *Logica statistică în gestiunea calității*).

În diferite colțuri din lume iau ființă nuclee puternice de cercetare în domeniul statisticii și aplicațiilor sale. Astfel, în R. D. Germană își desfășoară activitatea grupul condus de profesorul Egon Schindowski — pionier al aplicării controlului statistic al calității în Europa. Tot în R. D. Germană activează Institutul de Matematici Aplicate și Mecanică al Academiei de Științe a R. D. Germană (Institut für Angewandte Mathematik und Mechanik der D.A.W.) condus de profesorul Olaf Bunke — statistician de renume internațional, care este în același timp și redactorul șef al revistei „Cercetări Operaționale și Statistică Matematică" (Mathematische Operazionsforschung und Statistik). În Italia, profesorul Paolo Fortunati conduce un grup de cercetare la Universitatea din Bologna și editează revista „Statistica" cunoscută pe plan mondial sub numele de „Statistica din Bologna". La Universitatea Columbia din New York, profesorul Herbert Robbins, membru al Academiei de Științe al S.U.A. conduce departamentul de statistică al acestei celebre universități în care au lucrat mari titani ai statisticii ca W. A. Shewhart, A. Wald, S. Littauer și mulți alții. Acad. H. Robbins este deschizător de noi direcții de cercetare în statistică și este autorul (împreună cu R. Courant) unei lucrări de filozofie a matematicii intitulată *Ce este matematica?* (1944). Această lucrare a fost deja tradusă în mai multe limbi printre care în rusă, japoneză, germană etc. (ediția rusă: *Cito takoe matematika?*, Izdatelstvo Nauka, Moskva, 1947).

În Cehoslovacia, o amploare deosebită a căpătat-o studiul controlului proceselor tehnologice prin metode statistice. Statisticienii și inginerii cehoslovaci au rezultate valoroase în acest domeniu. Anežka Žaludová, Jana Jurečkova, O. Hanš, J. Hrábak, V. Horálek, J. Křepela, L. Kúbat, J. Likeš, L. Prouza, Z. Režný, A. Spaček, M. Ullrich, Fr. Zítek sînt numai cîteva nume din pleiada de statisticieni de renume ai școlii cehoslovace de statistică. (Unele realizări ale statisticienilor cehoslovaci sînt descrise, pe scurt, în lucrarea *A survey of some recent work in automatic statistical process control* de A. Žaludová, Z. Režný, M. Ullrich, și care este publicată în revista britanică „Journal of Applied Probability", vol. 5, 1968, pp. 43—54).

O activitate efervescentă pe plan mondial se desfășoară și în domeniul publicațiilor de specialitate. Cele mai vechi reviste de statistică au apărut în Marea Britanie. Odată cu creșterea rolului statisticii în viața modernă, acestea „se înmulțesc" în ritm accelerat, fiecare țară avînd chiar mai multe reviste de statistică. În 1929, apare „Giornale Dell'Istituto Italiano degli Attuari" (redactor F. P. Cantelli), în 1930, „The Annals of Mathematical Statistics" (azi cu două serii „Annals of Statistics" și „Annals of Probability"). În anul 1944 apare primul număr al revistei „Industrial Quality Control", care a jucat un rol deosebit în propagarea controlului statistic al calității, iar în 1959, revista „Technometrics" — dedicată în mod deosebit aplicațiilor statisticii în fizică, inginerie tehnologică și chimie.

„Teoria Veroiatnostei i eio Primenenia" (U.R.S.S.), „Aplikace Matematiky" (RSC), „Zastosowania Matematiky" și „Przeglad Statiszticzny" (RPP), „Trabajos de Estadistica" și „Cuadernos de Estadistica Aplicada" (Spania), Qualität und Zuverläsigkheit (RFG) etc., joacă azi un rol important în dezvoltarea statisticii matematice.

Trecînd pe alte continente, găsim „Sankhya" și „Journal of the Indian Statistical Association" în India, „Hinshitsu Kanri" (Controlul Statistic al Calității) și „Reports of Statistical Application Research" în Japonia,

„Acta Matematica Sinica" în R. P. Chineză, „Australian Journal of Statistics" în îndepărtata Australie etc.

Fiecare țară a dat lumii mari statisticieni, numele lor — dacă ar fi să le transcriem pe toate — ar ocupa poate un volum mai mare decît însăși cartea de față. Să menționăm doar pe cîțiva, pentru a ilustra varietatea colțurilor globului pămîntesc: Ahmed Sarhan (R.A. Egipt) — profesor la universitățile din Cairo și Alexandria, cu lucrări importante în domeniul repartiției exponențiale (vezi cap. III); G. M. el-Sayyad (Arabia Saudită), profesor la Universitatea din Riad, Károly Sarkadi (R.P. Ungară), cu lucrări fundamentale în mai multe domenii ale statisticii din care menționăm teoria testelor de concordanță (vezi cap. II) și controlul statistic al calității; Rafael Campo (Columbia) — tînăr om de știință cu rezultate valoroase în domeniul teoriei siguranței în funcționare a sistemelor și componentelor tehnice, Joao Saboia (Brazilia), de asemenea, cu lucrări importante în teoria siguranței; C. Radakrishna Rao (India) — deja un clasic al statisticii, rezultatele lui fiind nelipsite din cursurile de statistică de la orice facultate din lume; Des Raj (Nigeria), cu lucrări importante în domeniul teoriei selecției (vezi cap. II); Anders Hald (Danemarca) — specialist recunoscut în domeniul controlului statistic de recepție; Hassan O. Sirajdinov (U.R.S.S.) — profesor la Universitatea din Tașkent, savant de renume mondial, specialist, de asemenea, în controlul statistic de recepție; Tiago de Oliveira (Portugalia) — profesor la universitatea din Lisabona, specialist în teoria lanțurilor Markov și în controlul statistic al proceselor tehnologice etc. Cei dornici să-și completeze cunoștințele, nu numai cu nume, ci și cu informații concrete asupra lucrărilor elaborate de diferiți statisticieni, pot consulta așa-numitele reviste de referate internaționale.

Astfel, în U.R.S.S. apare „Referativnîi Jurnal" (seria pentru matematică), în S.U.A., „Mathematical Reviews", iar în R.D. Germană, „Zentralblatt für Mathematik". Toate aceste reviste de referate se găsesc la mai multe biblioteci din țara noastră (Biblioteca Academiei R.S. România, biblioteca Facultății de mate-

matică și mecanică a Universității, biblioteca Institutului Politehnic — pentru a nu cita decît cîteva biblioteci din București) și ele reprezintă un important instrument de lucru auxiliar în activitatea de cercetare și informare asupra tuturor domeniilor matematicii — nu numai asupra statisticii.

Pentru a întregi această scurtă privire asupra situației statisticii în lume, vom mai menționa că în ultimul timp, o importanță deosebită — alături de aspectele legate de calitatea produselor, a căpătat-o fiabilitatea acestora (siguranța lor în funcționare) care nu este altceva decît „dimensiunea în timp a calității" (Fl. Vasiliu).

Cercetătorul sovietic G. V. Drujinin sublinia într-o interesantă lucrare intitulată *Siguranța în funcționare a sistemelor* (Editura tehnică, București, 1968) că defecțiunile sistemelor tehnice apar de cele mai multe ori întîmplător, și, ca o consecință, modelarea comportamentului acestor sisteme nu poate fi decît probabilistico-statistică.

O intensă activitate de cercetare a început, în acest domeniu, în diferite țări din lume, formîndu-se adevărate grupe de cercetare, adevărate școli de cercetare. De o binemeritată faimă se bucură grupul condus de profesorul R. E. Barlow din Berkeley (S.U.A.), colectivul acad. B. V. Gnedenko (U.R.S.S), cel al omului de știință bulgar A. Obretenov, cel al profesorului polonez Szymon Firkowicz etc.

În fine, pentru a încheia, să amintim cîte ceva despre activitatea editorială în materie de cărți de statistică. Există o adevărată avalanșă pe plan mondial în acest domeniu. Înainte de cel de-al doilea război mondial — în afara lucrării lui Janko de care am pomenit, este notabilă lucrarea în două volume *Matematiceskaia Statistika (Statistica Matematică)* a marelui statistician sovietic V. I. Romanovski, apărută în 1938 la Tașkent.

Un an mai tîrziu, apare o altă lucrare fundamentală, a lui Shewhart, intitulată sugestiv *Metoda statistică din punctul de vedere al controlului calității (Statistical Method from the viewpoint of Quality Control*, 1939, editată cu asistența lui W. E. Deming). Ca apariție de bază se mai înscrie și lucrarea „titanului statisticii", R. A. Fisher, intitulată *Planificarea Experimentelor (The Design of Experiments*, 1935).

După cel de-al doilea război mondial, apariția cărților de statistică este efectiv explozivă. Pentru a da numai o indicație asupra acestui fapt, amintim că a apărut necesitatea întocmirii unor „Cărți despre cărțile de statistică" (vezi de exemplu: Kendall, M. G., Doig, A. G. *Bibliografia lucrărilor de statistică*, 3 volume, Oliver and Boyd, Edinburgh, 1968, și Menges, G., Leiner, B. *Bibliografia teoriei statistice a deciziilor*, Westdeutscher Verlag, Köln, 1968). Apariția în arena științifică mondială a țărilor în curs de dezvoltare a dus la îmbogățirea patrimoniului literaturii de specialitate și în domeniul statisticii.

Menționăm cîteva din editurile din diferite țări care au publicat numeroase cărți de statistică: Edición Revolucionaria (Instituto Cubano del Libro) din Havana, for de cultură recunoscut în întreaga lume de limbă spaniolă; Wydawnicztwa Naukowo-Techniczne din Varșovia, Akademiai Kiadó din Budapesta, John Wiley din New York (seria de statistică a apărut la inițiativa lui Shewhart și Wilks în 1954), Statistika din Moscova etc.

Iată — foarte pe scurt — numai cîteva din variatele aspecte ale activității în domeniul statisticii, care se desfășoară în lume. Altele vor fi prezentate pe parcurs, dar multe dintre ele — datorită vastității subiectului — sîntem nevoiți să le supunem atenției cititorilor indicînd bibliografia necesară.

Capitolul I

CE ESTE STATISTICA?

„Inferența statistică este acea ramură a metodei științifice care cu margini specificate de incertitudine exprimate în termeni probabilistici, face trecerea de la observații la concluzii privind populația".

G. Ciucu și V. Craiu [1]

Imaginea populară asupra statisticii nu este greu de ghicit: ceva care se ocupă cu studiul numerelor, cu așezarea lor în tabele, cu însumări pe linii sau pe coloane.

Această concepție nu este falsă: ea este însă profund incompletă și simplistă, iar apariția ei se poate explica destul de ușor.

Într-adevăr, „materia primă" a statisticii sînt numerele, iar obiceiul de a înregistra numere — de a colecta date statistice, cum spunem azi —, datează odată cu apariția societății omenești.

Administrațiile de stat s-au dovedit, de-a lungul timpului, cei mai mari „colecționari" de date statistice: asupra populației (recensămînturile), asupra comerțului (înregistrarea importurilor și exporturilor era făcută în Anglia, de exemplu, încă din secolul al XIII-lea), starea aprovizionării armatelor (în Roma antică se înregistrau soldații pe categorii, armamentul, uniformele etc.) și alte aspecte legate de societate în general.

În epoci mai recente, odată cu apariția presei deci și cu posibilitatea răspîndirii rapide a informațiilor, datele statistice au început să joace un rol important în ansamblul global al informațiilor oferite publicului larg.

[1] Ciucu, G., Craiu, V., *Inferență statistică*, Edit. didactică și pedagogică, București, 1974, p. 1.

Dezvoltarea societății a adus în prim plan noi tipuri de date statistice: asupra producției industriale, asupra situației economice, militare, financiare, asupra vremii, accidentelor de circulație etc. Alături de informația propriu-zisă, aceste date numerice au început să fie prezentate și sub formă de diagrame, grafice pentru a putea fi mai ușor receptate.

„Explozia informațională" din zilele noastre e însoțită, inevitabil, de date statistice asupra fenomenelor respective, încît iată deci că această concepție asupra statisticii a apărut în mod natural.

Astăzi, în majoritatea țărilor lumii există organisme specializate în strîngerea datelor statistice din toate domeniile de activitate care publică diferite materiale ca *Anuarele statistice* ale țărilor respective, Buletine statistice lunare etc.

Organizații internaționale ca F.A.O. (Organizația pentru alimentație și agricultură), W.H.O. (Organizația mondială a sănătății), U.N.E.S.C.O. etc. editează, la rîndul lor, publicații statistice în domeniul în care activează.

Acest prim pas, de colectare și prezentare a datelor statistice, este absolut necesar în cunoașterea unui fenomen luat în studiu. Dacă lucrurile s-ar fi oprit însă aici, cunoașterea fenomenului ar fi incompletă și deseori se pot trage concluzii false.

Acum intervine rolul metodelor specifice statisticii matematice de a prelucra datele culese, pentru a extrage maximum de informație asupra procesului respectiv. Evident, aceste metode nu pot face obiectul unei prezentări detaliate în presa zilnică: publicului larg i se prezintă, așa cum este normal, problema și concluzia la care s-a ajuns pe baza folosirii unor metode statistice.

Iată un exemplu: în orașul X, unde există mai multe fabrici de ciment, se constată o poluare a aerului în cartierele învecinate. În acest scop, s-a început o investigație pentru a vedea în cît timp nefolosirea unor filtre adecvate la coșurile de evacuare ar putea afecta compoziția aerului în așa grad încît aceasta să constituie un pericol pentru sănătatea locuitorilor.

Problema se poate prezenta în două moduri:

a°. Un mod tehnicist, cu detalii intermediare, cu descrierea unor metode statistice ce au fost aplicate;

b°. Un mod general-informativ de tipul: pe baza unor metode statistico-matematice s-a ajuns la concluzia că în aproximativ 2 ani, dacă nu se iau măsuri corespunzătoare de depoluare, concentrația de particule de ciment constatată în orașul X va atinge un grad neadmisibil.

În mod logic, modul b° este cel natural pentru publicul larg. Dacă vom începe să-i prezentăm problema scriind formule, interesul său, atît pentru informația în sine, cît și pentru statistica matematică, va scădea considerabil.

E adevărat că în modul b° de prezentare se ascunde un lung drum — adeseori destul de dificil, dar întotdeauna interesant — de la problema în sine pînă la o concluzie. În această parte stă munca cercetătorului care aplică metodele deja existente sau care, datorită problemei ce o studiază prin noutatea ei, creează noi metode de investigare.

Din informațiile pe care le primește zilnic prin intermediul diferitelor surse, cititorul observă însă un lucru foarte important și anume că statistica se aplică în cele mai neașteptate domenii, abstracție făcînd de acelea cu care el era deja obișnuit, că există *ceva* care deosebește statistica de celelalte științe cum ar fi fizica, chimia, economia etc.

Acest *ceva* este tocmai faptul că statistica nu are un obiect specificat, un cîmp limitat de acțiune: ea reprezintă un ansamblu de metode folosite pentru obținerea unor elemente noi de cunoaștere. Ea ajută celelalte științe să-și formuleze propriile lor legi.

Statisticianul indian B. N. Gupta consideră statistica drept „știința metodelor științifice și arta aplicării acestor metode". Într-adevăr, știința statistică este o parte componentă a metodei științifice, în general, tocmai pentru că ea urmărește obținerea de fapte noi, de informație nouă asupra procesului sau obiectului luat în studiu.

Statisticienii americani W. A. Wallis și H. V. Roberts au delimitat, foarte sugestiv, patru faze importante în investigația științifică și modul în care intervine statistica în decursul desfășurării lor.[1] Iată care sînt, în concepția lor, aceste faze:

1. *Observația:* se observă procesul sau obiectul considerat și se colectează faptele relevante pentru problema ce necesită rezolvarea;

2. *Formularea ipotezelor:* pentru explicarea faptelor observate se formulează unele ipoteze (sau teorii) asupra caracteristicilor detectate în faptele culese;

3. *Predicția:* pe baza ipotezelor formulate se fac anumite deducții. Dacă teoria avansată se dovedește satisfăcătoare, atunci aceste deducții constituie elemente noi de cunoaștere ce nu au fost deduse pe cale experimentală, ci pe baza teoriei formulate. Aceste noi elemente de cunoaștere se numesc *predicții* și ele pot anticipa (cu un oarecare risc) anumite caracteristici în studiile viitoare asupra unui proces de același tip sau de tip asemănător;

4. *Verificarea:* se colectează noi fapte pentru a testa predicțiile efectuate pe baza teoriei formulate anterior.

Se observă că avem aici de-a face cu un ciclu care se repetă, însă scopul cu care se face colectarea faptelor aici este diferit față de pasul inițial. Dacă teoria nu se verifică — ceea ce se constată prin apariția unor predicții ce o contrazic — atunci se reformulează ipotezele și ciclul continuă.

În această prezentare, se poate recunoaște aplicarea cunoscutei scheme pusă în evidență de Lenin: Practică—Teorie—Practică, în care practica este criteriul fundamental de cunoaștere al adevărului [2].

---

[1] Wallis, H. A., Roberts, H. V., *The Nature of Statistics*, Collier Books, New York, 1962.

[2] Lenin, V. I., *Caiete filozofice*, ESPLP, București, 1956, p. 176.

Să facem aici o scurtă paranteză și să subliniem faptul că gîndirea materialist-dialectică este îmbrățișată de majoritatea oamenilor de știință ce lucrează în domeniul statisticii. S-a întîmplat ceea ce clasicii materialismului dialectic au prevăzut cu multă vreme în urmă: dezvoltarea științei nu va face decît să îmbogățească cu noi elemente concepția filozofică a materialismului dialectic, demonstrînd odată în plus viabilitatea sa.

Vorbind despre metoda științifică, statisticianul indian S. K. Ekambaram spunea[1]:

„Cînd observațiile experimentale sînt capabile să verifice și să confirme consecințele unei ipoteze, atunci sîntem în posesia unei legi științifice. *O afirmație ce nu poate fi experimental verificată, nu poate fi o lege științifică*" (s.n.).

Revenind la periodizarea Wallis-Roberts, să urmărim modul în care statistica intervine în investigația științifică.

În prima etapă, statistica sugerează cum să se efectueze observațiile și cum să se interpreteze. O contribuție importantă a statisticii în acest pas al investigației constă și în selectivitatea impusă observației care are drept consecință imediată economia de probe și experimente.

În etapa a doua, statistica este utilă prin sintetizarea, clasificarea și prezentarea datelor culese astfel încît să se sugereze avansarea ipotezelor cele mai plauzibile.

În etapa a treia, statistica intervine de astă dată prin metode specifice, matematice de prelucrare a informației și efectuare a predicțiilor.

Astfel, dacă s-a stabilit, de exemplu, într-un experiment că elasticitatea (Y) unui anumit material sintetic depinde liniar de temperatura mediului ambiant

[1] Ekambaram, S. K., *The Statistical Basis of Acceptance Sampling — An Introduction for Industrial Executives*, Asia Publishing House, Calcutta, 1960, p. 21.

(T), atunci pe baza unor metode statistico-matematice se determină parametrii ecuației liniare respective:

$$Y = \hat{\alpha}T + \hat{\beta} \qquad (\hat{\alpha}, \hat{\beta} \text{ — parametri, determinați numeric}) \quad (1)$$

În baza acestei ecuații, se poate efectua o predicție a elasticității materialului studiat la o temperatură specifică unei anumite condiții și astfel, dacă se obține elasticitatea scontată, se poate întrebuința materialul respectiv într-o condiție experimentală ce se va crea ulterior (decizia).

În etapa a patra, statistica este deosebit de utilă în verificarea ipotezelor avansate, prin determinarea faptului dacă noile date observate sînt în concordanță sau nu cu predicțiile respective.

Aici trebuie să ne oprim puțin și să discutăm un aspect foarte important: chiar dacă teoria este adevărată, verificarea predicțiilor nu va furniza exact același rezultat cu cel dedus teoretic.

Diferențele — care în cazul că teoria este adevărată sînt nesemnificative —, pot apărea datorită unor cauze inerente, cum ar fi eroarea experimentală sau variabilitatea naturală, intrinsecă a materialului statistic, adică a datelor de observație. Dacă diferențele sînt mari și ele nu pot fi atribuite unor cauze inerente, atunci trebuie determinată natura acestor cauze care, în esență, vor conduce la modificarea teoriei, la avansarea altor ipoteze asupra fenomenului considerat.

Ceea ce trebuie să reținem atunci cînd începem să gîndim statistic este faptul că în natură, în activitatea productivă sau științifică există această *variabilitate naturală a fenomenelor, a proceselor* care deseori vine în contradicție cu vechile noastre concepții despre rigurozitatea științifică.

În realitate, adevărata rigurozitate trebuie să țină cont de fenomenul real.

Să considerăm cîteva exemple pornind de la faptele cele mai simple, aparent banale, din experiența noastră zilnică.

Cumpărați o cutie cu ace de gămălie și răsfirați-le pe o masă. Veți observa că unele sînt mai scurte, altele mai lungi (diferență evident sesizabilă dacă privim cu atenție), că unele sînt mai groase iar altele mai subțiri, că unele au vîrful mai ascuțit, iar altele mai puțin ascuțit. Desigur, dacă vom continua examinarea lor vom găsi probabil și alte deosebiri. În esență, ceea ce vrem să subliniem este că foarte rar vom găsi două exemplare perfect identice deși este de la sine înțeles că toate acele au fost fabricate după același procedeu, din același tip de materie primă etc.

Avem aici de-a face cu variabilitatea naturală a unui proces tehnologic. Nu numai temperatura mediului ambiant, variațiile de curent electric etc., dar și schimburile de lucru contribuie la această variabilitate naturală a procesului, care se traduce, de fapt, prin variabilitatea observată în datele experimentale privind caracteristicile de calitate ale produselor.

Modul de gîndire determinist (nu este greșit dacă întrebuințăm și termenul de fixist) ne-a obișnuit cu ideea că dacă avem de-a face cu un anume obiect — să zicem un aparat de radio de un tip specificat — atunci toate aparatele de același tip trebuie să aibă absolut aceleași caracteristici funcționale, estetice și de durată în funcționare.

Propria noastră experiență ne-a învățat că acest mod de a privi lucrurile este nerealist. În timp ce aparatul nostru de radio a funcționat foarte bine timp de 10 ani, după care, defectîndu-se, a trebuit să-l înlocuim, aparatul unui cunoscut, de același tip și cumpărat în aceeași zi, continuă să refuze să se defecteze!

Dacă vom căuta, vom găsi exemple nenumărate de acest gen. Variabilitatea statistică este unul din exemplele mișcării dialectice, caracteristice lumii materiale, care are un caracter universal și etern.

1.1. DOMENIILE STATISTICII

„Cu fiecare zi viața este influențată, din ce în ce mai mult, de deciziile bazate pe informații cantitative".

V. Craiu [1]

Ca orice altă știință, este de bănuit că și statistica are multe domenii specifice. Totuși, ca și în cazul chimiei unde se poate face o împărțire generală în chimie anorganică și chimie organică, și statistica poate fi clasificată în două mari categorii și anume:

— statistica descriptivă;

— statistica analitică.

Conținutul fiecărei părți nu este greu de lămurit și este sugerat de însăși denumirea lor.

Statistica descriptivă se ocupă cu clasificarea, prezentarea și sintetizarea datelor de observație. Este deci partea din știința statisticii care intervine cu deosebire în etapa a doua a investigației științifice din schema Wallis-Roberts. Totodată, statistica descriptivă concentrează informația existentă în datele respective cu ajutorul anumitor *indicatori statistici* care, în fond, sînt niște numere ce exprimă caracteristici sau tendințe ale fenomenului investigat.

Cel mai răspîndit indicator statistic, cu care sîntem obișnuiți din presa zilnică, din emisiunile de radio și televiziune este *media* (sau *valoarea medie*). Astfel, ne sînt familiare noțiuni ca „venitul mediu pe cap de locuitor" dintr-o anumită țară, cu „cantitatea medie de precipitații" căzută într-o lună într-o anumită zonă geografică etc.

Desigur, media este un indicator foarte important, el exprimînd *tendința centrală* a datelor de observație dintr-un proces sau fenomen luat în studiu. Dar media singură nu este suficientă pentru a căpăta o imagine

[1] Craiu, V. *Verificarea ipotezelor statistice*, Editura didactică și pedagogică, București, 1972, p. 3.

reală. Astfel, este suficient să dăm un exemplu în care numărul 10 se poate obține ca medie a numerelor 10, 10, 10, 10, dar în același timp ca medie a numerelor 1, 1, 1, 37, pentru a vedea că împreună cu media mai sînt necesari și alți indicatori care să aducă elemente complementare ca: împrăștierea datelor, valoarea cea mai frecventă (care apare cel mai des), valoarea maximă etc.

Este tocmai rolul statisticii descriptive să prezinte toți acești indicatori alături de tabele, reprezentări grafice, diagrame care să poată sugera apoi formularea unor ipoteze asupra problemei considerate.

Există multe lucrări de statistică descriptivă care tratează în detaliu toate aceste elemente [1].

Cealaltă parte a statisticii — *statistica analitică* — folosește metode specifice, matematice pentru extragerea informației din materialul statistic organizat de metodele descriptive. Ea intervine deci în etapa a treia din schema Wallis-Roberts.

Rolul metodelor analitice este să dea, în ultimă instanță, un răspuns la o anumită problemă concretă: este sortimentul A de fibre sintetice mai rezistent decît sortimentul B? Schimbul de noapte și schimbul de zi dintr-o anumită întreprindere produc obiecte a căror calitate diferă semnificativ? Două utilaje de același tip, dar produse de fabrici diferite au în medie aceeași durată de funcționare?

Nu s-au enunțat aici decît cîteva probleme *de comparare*, cu importanță practică evidentă a căror rezolvare se poate face prin metode statistice. Ceea ce interesează este, în mod normal, concluzia — care poate fi sintetizată într-un singur cuvînt DA sau NU —, indiferent cît de complicată ar fi metoda statistică ce a dus la unul din răspunsuri. Detaliile matematice, de calcul etc. nu mai prezintă importanță, iar răspunsul este util *chiar dacă este negativ*.

[1] De exemplu, vezi Ciucu, G., Craiu, V., *Introducere în teoria probabilităților și statistica matematică*, Edit. didactică și pedagogică, București, 1971.

1.2. CÎTEVA ELEMENTE DE STATISTICĂ DESCRIPTIVĂ

„... nici o metodă statistică prin ea însăși nu poate să împiedice greșala, inexactitatea, raționamentul fals sau concluziile eronate. Datele originale trebuie să fie exacte, metodele trebuie adecvat aplicate și rezultatele trebuie interpretate de cineva care cunoaște bine nu numai metodele, ci și domeniul în care ele sînt aplicate".

V. Craiu [1]

Orice știință operează cu cîteva noțiuni fundamentale, celelalte fiind derivate directe sau indirecte ale acestora.

În statistică, noțiunea de bază este aceea de *colectivitate* sau *populație*, termen prin care se poate desemna orice mulțime de elemente luate în studiu. Denumirea de „populație" s-a păstrat din timpurile în care statistica se ocupa cu precădere de populații în sensul propriu al cuvîntului.

Astăzi însă el se folosește pentru orice grup de elemente: piese de un anume tip produse de o anume mașină-unealtă, locuitori dintr-un anumit sat, cazuri de îmbolnăviri cu un anumit virus etc.

O particularitate specifică statisticii pe care cititorul a remarcat-o, credem, pînă acum este aceea că statistica nu se ocupă cu un element (individ) luat ca atare, ci cu colectivități, cu grupuri de elemente ce posedă o anumită trăsătură comună. Această trăsătură comună se numește *caracteristică*.

Caracteristicile sînt, după natura lor, de două categorii:

a°. *Calitative* — ce au drept particularitate faptul că nu se pot exprima numeric. De exemplu: culoarea

[1] Craiu, V., *Verificarea ipotezelor statistice*, Editura didactică și pedagogică, București, 1972, p. 4.

ochilor, concepțiile, starea operațională a unui produs (defect, în stare funcțională) etc.

De observat că aceste caracteristici sînt descrise, de obicei, de adjective sau adverbe, desemnînd, în general, un anumit *atribut*: de a fi corespunzător sau nu (pentru un produs), de a fi verde, albastru, negru (pentru o culoare de ochi, de pildă) etc. De aceea, caracteristicile calitative se mai numesc și caracteristici atributive sau mai simplu *atribute*.

Totuși, și în cazul acestor caracteristici se poate obține o descriere numerică. De exemplu, să presupunem că într-un colectiv de elevi acordăm 5 puncte elevilor care au notele la toate materiile peste 8 inclusiv. Apoi, mai acordăm 5 puncte celor care pot să poarte o conversație fluentă într-una din limbile de mare circulație. Mai acordăm încă 5 puncte celor care sînt colaboratori la una din revistele pentru tineret („Gazeta Matematică", „Revista de Fizică și Chimie" sau „Limba și Literatura română") etc.

Acest procedeu reprezintă o *codificare* a unor atribute și servește, într-un fel, drept criteriu de clasament.

Este însă destul de limpede că această metodă nu îndreptățește deloc folosirea operațiilor uzuale cu numere; într-adevăr, a aduna, de exemplu, punctajele tuturor elevilor și a împărți rezultatul la numărul lor — pentru a afla astfel *o medie* — este lipsit de sens deoarece numărul obținut nu va conține nici o informație; că așa stau lucrurile este foarte simplu de constatat imaginînd, de pildă, că s-a obținut ca medie numărul 10.

Ce concluzie putem trage? Că în medie:

1°. Elevii noștri au note mai mari decît 8 la toate materiile și conversează fluent într-o limbă de mare circulație;

2°. Elevii noștri au note mai mari decît 8 la toate materiile și sînt colaboratori la o revistă pentru tineret;

3°. Elevii noștri conversează fluent într-o limbă de mare circulație și sînt colaboratori la o revistă pentru tineret.

Desigur, oricare dintre aceste concluzii este falsă. Ne-o spune însăși intuiția (care în statistică nu trebuie niciodată disprețuită!) dar în mod sigur ne-o va spune un profesor real dintr-o clasă reală, sau un elev real dintr-o clasă la fel de reală.

Ceea ce însă putem obține cu o astfel de codificare a caracteristicilor atributive, este o informație de tipul:

Ionescu și Popescu au totalizat 15 puncte;

Vasilescu, Georgescu, Grigorescu au totalizat 10 puncte ș.a.

Sau, neglijînd numele (statistica e ingrată în acest sens!), putem forma următorul tabel:

Nr. crt.	Nr. de puncte	Nr. de elevi
1	15	2
2	10	3
3	5	15

În această formă, informația obținută este statistică, deoarece pentru o anumită caracteristică atributivă am numărat indivizii ce au întrunit-o.

Exemplul ales conține, ascuns, un viciu, care însă nu modifică deloc afirmațiile cu privire la atribute, codificare etc. Am introdus în mod deliberat acest viciu pentru a sublinia cît este de important în statistică modul de abordare al unei investigații. Să revenim astfel la tabelul de mai sus: la nr. 3, 5 puncte au fost întrunite de 15 elevi. Care însă din caracteristicile atributive menționate sînt îndeplinite? Desigur, se poate spune că *cel puțin una*. Dar care?

Nu e complicat să ne dăm seama că informația s-a diluat datorită modului în care am făcut codificarea. Pentru fiecare atribut trebuia să fi acordat *alt punctaj*.

Încercați să refaceți acest exemplu acordînd, de pildă, cîte 5, 3, 2 puncte respectiv pentru fiecare caracteristică atributivă considerată. Ca număr de elevi

luați chiar numărul real de colegi din clasă (sau grupa) dumneavoastră. Dacă nu se potrivesc caracteristicile atributive din exemplul nostru, căutați altele.

Va fi poate prima „anchetă statistică" pe care o faceți.

Nu veți obține nici un rezultat spectaculos; dar drumul pînă la aplicarea metodelor statistice în domeniul în care veți lucra este lung și nu e deloc ușor. Un vechi proverb arab spune: „Cine vrea să facă cu adevărat un lucru, găsește pînă la urmă mijloacele să-l facă. Cine nu vrea, găsește întotdeauna o scuză".

b°. *Cantitative* — sînt cele care pentru fiecare element se pot exprima numeric, în funcție de unitatea de măsură adecvată.

Astfel, rezistența la rupere a unor fire relon, venitul lunar al lucrătorilor dintr-o anumită întreprindere, numărul de descendenți al locuitorilor dintr-un anumit sat etc., constituie cîteva exemple de caracteristici cantitative.

Nu există domeniu de activitate umană în care să nu intervină caracteristici cantitative. Sfera economico-productivă furnizează cu precădere aceste caracteristici pe care putem să le considerăm materia primă de bază a statisticii.

Caracteristicile cantitative mai poartă numele și de *variabile*. Cu această denumire sîntem familiarizați mai de mult, atît din cărțile de școală cît și din viața noastră de zi cu zi. Presiunea atmosferică, cantitatea de precipitații din orașul în care trăim, înălțimea sau greutatea proprie, etc. sînt exemple de variabile.

N-am ales aceste exemple cu totul întîmplător. Dacă le examinăm cu atenție, vom distinge, după natura lor, două tipuri de variabile și anume:

— *cele discrete:* care pot lua doar un număr discret de valori. De pildă, între grupurile de persoane care au doi frați și cele care au trei nu mai există nici o valoare intermediară. Trecerea se face deci direct de la o valoare la alta;

— *cele continue:* care pot lua orice valoare dintr-un anumit interval.

Iată, între un individ cu o înălțime de 1,70 m și altul cu o înălțime de 1,71 m există, din punct de vedere teoretic, un număr infinit de cazuri, deci înălțimea unei persoane este o variabilă continuă.

Totuși, cred că n-ați întîlnit niciodată pe cineva care întrebat ce înălțime are să răspundă: „un metru, șaptezeci de centimetri, opt milimetri și 6 microni". Acest lucru nu se va întîmpla, fie că precizia aparatelor cu care efectuăm măsurătorile este limitată, fie că o precizie exagerată nu este în toate cazurile folositoare pentru ceea ce urmărim în investigația noastră.

Astfel, măsurătorile sau datele de observație se grupează în cadrul unei anumite unități și deci din punct de vedere practic se lucrează cu forma discretă chiar dacă variabilele sînt continue.

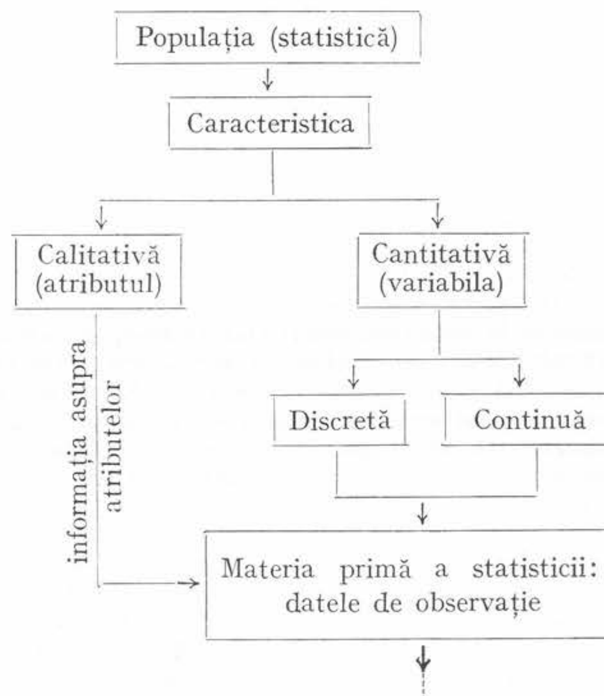
Aici trebuie să lămurim un lucru care deseori dă naștere la confuzii: mulți sînt înclinați să creadă că variabilele discrete trebuie să ia neapărat numai valori întregi și că numerele fracționare sînt tipice numai pentru variabilele continue, cu alte cuvinte diferența dintre continuu și discret este confundată cu diferența între măsurătorile cu numere întregi și cele cu numere fracționare.

Acest lucru este fals și se poate arăta foarte ușor considerînd o caracteristică cantitativă ce poate lua *numai* valorile: 1,053; 1,065, 1,076, 1,089. Cu toate că avem trei zecimale, variabila este discontinuă, deoarece trecerea de la o valoare la alta se face fără vreo valoare intermediară.

Exemplul nostru este evident teoretic: poate o caracteristică de acest gen nici nu există. Aceasta însă nu afectează raționamentul de mai sus. Încercați să găsiți o caracteristică reală ale cărei valori se măsoară prin numere „cu zecimale", dar care caracteristică să aibă natura discontinuă a celei din exemplul „construit". Aveți la dispoziție fizica, chimia, biologia sau orice altă știință care vă place mai mult. Faceți practică în producție: la întreprinderea la care vă duceți căutați să observați în procesele tehnologice respective ce caracteristici intervin. Chiar dacă nu veți găsi caracteristica dorită, veți găsi altele și veți veni în contact

direct cu variabilitatea statistică de care vorbeam înainte.

Tot ce am discutat pînă acum, poate fi sintetizat în următoarea schemă simplă:



Din schemă — pe care o vom completa pe măsură ce înaintăm în subiect —, deducem că ceea ce ne interesează este populația statistică luată în considerare. Nu am văzut încă prin ce modalități reușim să extragem informația asupra populației respective.

Tot intuiția ne va spune că numai în foarte puține cazuri vom putea extrage informația de la fiecare element al populației luat în parte. Multe lucruri pledează împotriva unui astfel de mod de investigație. În primul rînd, în multe situații, acest procedeu — numit în statistică *enumerare completă* — este practic imposibil

de realizat. Iată, să ne gîndim la o fabrică de conserve care livrează unui beneficiar o anumită cantitate de conserve.

La primire, beneficiarul — deși presupunem că are încredere nelimitată în producător — trebuie să se asigure într-un fel de calitatea produselor primite. Din această cauză, el sacrifică o parte din conservele primite adică le deschide și le testează anumite caracteristici specifice: cantitatea, prospețimea, ingredientele etc. Evident, dacă se vor deschide toate cutiile primite atunci beneficiarul (care, de obicei, în acest circuit economic este o întreprindere de comerț interior sau exterior) nu va mai avea ce vinde.

Am ales aici un exemplu tipic din controlul calității produselor ce implică distrugerea produselor în vederea cercetării calităților lor.

Pe de altă parte, problemele de organizare, *de cost*, de timp fac, de asemenea, nerentabilă enumerarea completă. Din acest motiv, extragerea informației se face doar dintr-o submulțime a populației luate în studiu, submulțime ce poartă numele de „selecție" (unii autori preferă termenul de „sondaj").

Cercetarea astfel efectuată este deci *o cercetare* selectivă. Procedeul nu este deloc nou pentru noi. Ne-am întîlnit deseori cu el în paginile ziarelor la rubricile de anchetă socială. Dacă ne reamintim rubrica din „România liberă", „Vrem să facem un test", atunci vom regăsi imediat exemple de investigație selectivă. De pildă, cînd s-a dorit o imagine asupra stării de igienă a unor unități de alimentație publică din orașul București, din populația tuturor acestor unități s-a ales — după anumite criterii — doar o parte din acestea.

Cineva ar putea spune însă că, în acest caz, enumerarea completă este posibilă, deoarece numărul unităților de alimentație publică dintr-un oraș, oricît ar fi el de mare, nu este de ordinul sutelor de mii sau milioanelor.

Observația este adevărată, dar chiar și în situațiile în care enumerarea completă este posibilă (atît teo-

retic cît și practic) folosirea metodei selective este de preferat deoarece:

1°. scurtează timpul necesar obținerii datelor și prelucrării lor;

2°. permite un studiu mai aprofundat datorită concentrării atenției asupra unei părți din populație;

3°. reduce costul cercetării — problemă deosebit de actuală în toate domeniile.

Ați remarcat, desigur, în exemplul de mai sus că alegerea unităților de alimentație publică a fost făcută într-un anumit fel. Necesitatea unor criterii în alegerea selecției este indiscutabilă: dacă reporterul ar fi optat, de pildă, doar pentru restaurantele din cartierul său, atunci oricine ar fi putut să tragă concluzia că rezultatele anchetei nu se pot extinde pentru întreaga rețea din oraș, deoarece ceea ce se petrece într-un anumit cartier nu este întotdeauna reprezentativ pentru un centru urban de largă întindere.

Din această cauză, ne dăm seama că alegerea trebuie făcută *la întîmplare* adică să acordăm elementelor populației o șansă (posibilitate) egală de a fi introduse în selecție. Ceea ce dorim, de fapt, este să obținem o selecție cît mai *reprezentativă*, adică structura ei să difere cît mai puțin de aceea a populației respective.

Identitatea deplină a structurii selecției nu este imposibil de obținut, dar ea se realizează practic cu totul întîmplător.

La reprezentativitate mai contribuie și numărul elementelor selectate (sau volumul selecției, în limbaj statistic). Dacă numărul este prea mic, s-ar putea ca selecția să fie nereprezentativă. Pe de altă parte, dacă el este prea mare, intervin dezavantajele de ordin economic pe care le-am discutat la enumerarea completă.

Iată deci o problemă de mare importanță în investigația statistică: determinarea volumului de selecție optim. Aici cuvîntul optim are sensul de minim: volumul de selecție minim necesar obținerii unei informații maxime pentru minimul respectiv. Într-adevăr, pare lipsit de sens (și chiar este) să folosim mai multe elemente ale selecției — deci să cheltuim mai mult — dacă nu este necesar, adică dacă nu obținem nici o informație

importantă în plus care să lărgească sfera cunoașterii noastre despre un anumit fenomen sau proces.

Să ne oprim acum asupra unui alt aspect: lucrînd cu o selecție deci cu o parte a unui întreg, este inevitabil ca rezultatele investigației statistice să fie afectate de erori. Ceea ce realizează metodele statistice este faptul de a prevedea aceste erori, deci de a efectua cercetarea în deplină cunoștință de cauză. Este poate unul din meritele cele mai mari ale științei statistice.

1.3. PREZENTAREA DATELOR STATISTICE

„... tehnicile statistice nu pot crea nimic în plus față de starea inițială. Datele culese dintr-un experiment prost planificat vor oferi o falsă informație care în mod sigur conduce la o acțiune eronată".

Takeshi Suzuki [1]

Înainte de a considera mai amănunțit această problemă, sîntem nevoiți să facem iar o paranteză pentru a lămuri un aspect care deseori este trecut cu vederea.

Desigur, toți citim presa zilnică, reviste periodice, citim cărți, vizionăm diferite emisiuni de televiziune. Am văzut deci sau știm, fără chiar să fi citit vreo lucrare specială din domeniul statisticii, că datele statistice sînt prezentate sub formă de tabele, grafice, diagrame etc. Știm mai puțin, poate, despre modul lor de alcătuire, despre clasificarea lor și așa mai departe. Cu ceea ce venim noi în contact sînt numerele (datele) ce reprezintă anumite proprietăți.

În spatele acestor numere — pe lîngă fenomenul sau procesul concret pe care îl reprezintă —, se ascunde

[1] S u z u k i, T., *Statistical Quality Control as a tool for higher productivity*. În: „Symposium on Quality Control", A.P.O., Tokyo, 1964, p. 155.

și un alt proces și anume *procesul de măsurare*. Majoritatea datelor statistice reprezintă, în fond, niște rezultate ale unor măsurători: diametre ale unor piese cilindrice, greutăți ale unor indivizi, tensiuni într-o rețea electrică, producții de grîu la hectar etc.

Noțiunea de măsurătoare se utilizează și într-un sens mult mai general decît în exemplele de mai sus. Astfel, se vorbește de a măsura succesul de librărie al unei cărți de beletristică, popularitatea unui actor etc.

Ceea ce ne interesează pe noi sînt măsurătorile obținute ca urmare a existenței următoarelor trei elemente:

1°. un sistem fizic sau un obiect oarecare;

2°. o proprietate fizică sau chimică;

3°. o metodă de determinare a valorii (sau valorilor) acestei proprietăți.

Acest tip de măsurători acoperă majoritatea cazurilor ce le întîlnim în practică.

Într-un anume sens, procesul de măsurare poate fi comparat cu un proces de producție, ce are drept rezultat producerea unui obiect, în timp ce procesul de măsurare „produce" un număr sau o mulțime de numere.

Condițiile de mediu — ca temperatura ambiantă, presiunea, tensiunea curentului electric etc. — afectează mai mult sau mai puțin caracteristicile calitative ale obiectelor produse. În același timp, condițiile de mediu afectează și procesul de măsurare, producînd variații sesizabile în numerele rezultate din operațiile de măsurare respective. Variațiile constituie *eroarea experimentală* pe care am amintit-o și în paragraful precedent.

Eroarea experimentală nu trebuie confundată cu eroarea în sensul propriu al cuvîntului adică cu greșeala. Greșelile rezultă, de cele mai multe ori, din neaplicarea corectă a procedeelor de măsurare, înregistrarea incorectă a rezultatelor, neatenție etc.

În domeniul social, datorită dificultății unor probleme specifice (cum ar fi, de exemplu, înregistrarea exactă a vorbitorilor unei anumite limbi) sau, pur și simplu, superficialității cu care sînt conduse unele investigații statistice, apar deseori date ce diferă enorm de la sursă la sursă.

Un exemplu: două almanahuri americane din același an (1950) și anume „World Almanac" și „Information Please Almanac" treceau ca vorbitori ai limbii române următoarele cifre:

W.A.	19 400 000,
I.P.A.	16 000 000,

ca să nu mai amintim de alte diferențe și mai mari (de pildă, la vorbitorii de limbă spaniolă o diferență de ... 30 de milioane!).

Apare deci foarte clară importanța acurateței datelor pe care vrem să le prelucrăm statistic. Dacă ele nu sînt reale, atunci degeaba vom aplica cele mai sofisticate metode statistice: informația ce o vom obține va fi falsă, iar decizia ce o luăm pe baza unei informații false este întotdeauna împotriva propriilor noastre interese.

Revenind la măsurători, să remarcăm că măsurătorile de tipul pe care le-am considerat mai înainte implică de multe ori o relație de genul celor exprimate cu ajutorul unei funcții.

Să considerăm un exemplu: Se știe că rezistența unui conductor depinde de lungimea conductorului (l), de secțiunea sa (S) și de natura substanței din care este construit, natură exprimată printr-un factor multiplicativ (ρ) numit rezistivitatea materialului.

Astfel:

$$R = \rho \frac{l}{S} \quad \text{(unitate de măsură: ohmi } (\Omega))$$

Cunoscînd deci rezistivitatea, măsurarea rezistenței unui conductor se poate reduce la măsurarea unei lungimi și a unei secțiuni.

Natura cantităților R și l, S este evident diferită, dar putem afirma că am măsurat rezistența chiar dacă, în realitate, am măsurat o lungime și o suprafață. Aceasta este posibil tocmai datorită relației funcționale de mai sus.

Nu mai este necesar să spunem că există foarte multe exemple în care să intervină o relație funcțională.

Prin urmare, în procesul măsurării intervin trei elemente:

a°. proprietatea ce trebuie determinată (P);
b°. cantitatea sau caracteristica măsurată (M);
c°. legătura (relația funcțională) dintre aceste două elemente.

Grafic, aceasta se poate imagina astfel:

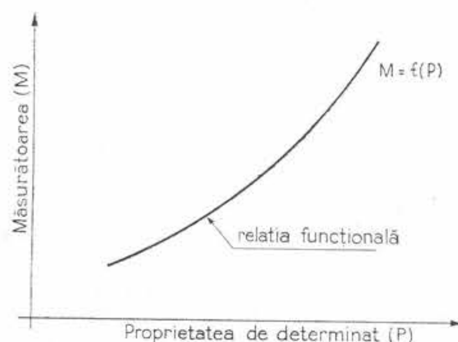


Fig. 1 — Legătura între cele trei elemente ale unui proces de măsurare

Ar fi interesant să concretizăm mai mult descrierea procesului de măsurare: în primul rînd să observăm că în anumite situații practice, simbolul P rămîne mai mult sau mai puțin la nivelul de simbol. De exemplu, limpezimea unui lichid, netezimea unui cristal etc. În astfel de situații, relația funcțională de la punctul c° rămîne necunoscută.

În al doilea rînd, dacă privim cu atenție figura 1 observăm că am ales pentru „f" o funcție monotonă, în cazul de față crescătoare.

Dacă funcția nu este monotonă, adică dacă arată ca în fig. 2, atunci pentru două valori diferite ale lui P (P_1 și P_2) se obține aceeași măsură M, fapt care creează o situație de incertitudine: care, în realitate, trebuie considerată ca fiind reprezentată de M?

O altă problemă este următoarea: numărul care reprezintă măsurătoarea este unic?

O bară de metal de 1 m poate foarte bine fi considerată ca o bară de metal de 100 cm. Temperatura aerului la un anumit moment și într-un anume loc poate avea 32 grade Fahrenheit, dar tot așa de bine zero grade Celsius etc.

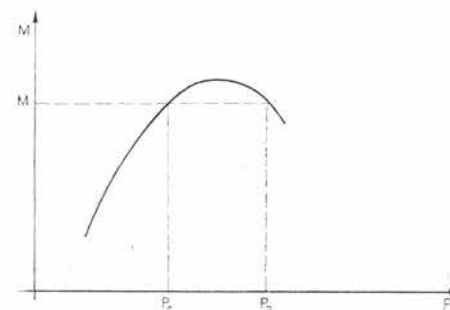


Fig. 2 — Funcție nemonotonă asociată unui proces de măsurare

Numerele, deși exprimă aceeași proprietate, sînt diferite. Acest lucru este posibil datorită așa-numitelor *transformări de scară*. Putem întrebuința, în măsurare sau în înregistrarea datelor, orice scară, cu condiția să o folosim numai pe cea aleasă inițial în tot cursul cercetării.

Astfel, pentru temperaturi, în Europa continentală se folosesc gradele Celsius, iar în S.U.A. și Canada, de exemplu, gradele Fahrenheit.

Trecerea de la o scară la alta se face, de obicei, printr-un factor multiplicativ (de exemplu $1\ \text{m} = 10 \cdot 10\ \text{cm}$) sau printr-o relație liniară (sau, în unele cazuri, neliniară).

De exemplu, legătura între gradele Celsius și Fahrenheit este dată de relația:

$$T_c = \frac{1}{9}[5 \cdot T_F - 160] \qquad (2)$$

care este o relație liniară, (astfel, 32°F = 0°C).

Uneori, transformările de scară schimbă forma relației inițiale. De exemplu, capacitatea calorică (q) care este dată de o relație de tip multiplicativ:

$$q = m \cdot c \quad \text{(unități: joule/grad)} \qquad (3)$$

(să ne reamintim: capacitatea calorică este cantitatea de căldură care trebuie transmisă unui corp de masă m pentru a-i mări temperatura cu un grad; simbolul c este așa-numita căldură specifică, cantitatea de căldură ce trebuie transmisă unității de masă dintr-un corp pentru a-i mări temperatura cu un grad), devine o relație de tip aditiv, dacă se aplică logaritmii:

$$\log q = \log m + \log c. \qquad (4)$$

În aceeași situație sînt și multe alte legi ale fizicii:
— legea *Boyle-Mariotte*: $p \cdot V = \text{const.}$,
— legea *Mendeleev-Clapeyron*: $p \cdot V = R \cdot T$ etc.

Desigur, aceste transformări nu sînt utile în toate cazurile; în unele, ele nici nu se pot aplica. Dar există multe situații întîlnite în practică, pe care le vom discuta mai tîrziu, în care transformările logaritmice nu numai că sînt utile, dar reprezintă și calea cea mai simplă în investigația statistică.

Să ne oprim aici cu considerațiile despre măsurători și procesele de măsurare, ceea ce nu înseamnă că am epuizat problema. De fapt, în acest domeniu avem de-a face cu o întreagă disciplină cunoscută azi sub numele de *Metrologie*, ale cărei tradiții în țara noastră sînt destul de vechi. În acest sens, cititorul are la dispoziție întreaga colecție a revistei „Calitatea Producției și Metrologiei“, revistă ce apare sub egida Institutului central de metrologie și a Inspectoratului de stat pentru controlul calității produselor din București. Aici se pot găsi lucrări interesante privind practica statistică în diferite ramuri industriale, probleme specifice ale proceselor de măsurare, traduceri de profil din literatura străină etc.

În schema prezentată la capitolul I/1.2, și pe care am promis că o vom completa pe parcurs, a apărut deci un nou element: între caracteristicile cantitative (dis-

crete sau continue) și materia primă a statisticii (datele de observație sau selecția, sau încă datele experimentale), se interpune procesul de măsurare, selecția (vom prefera acest termen) fiind rodul direct al acestui proces.

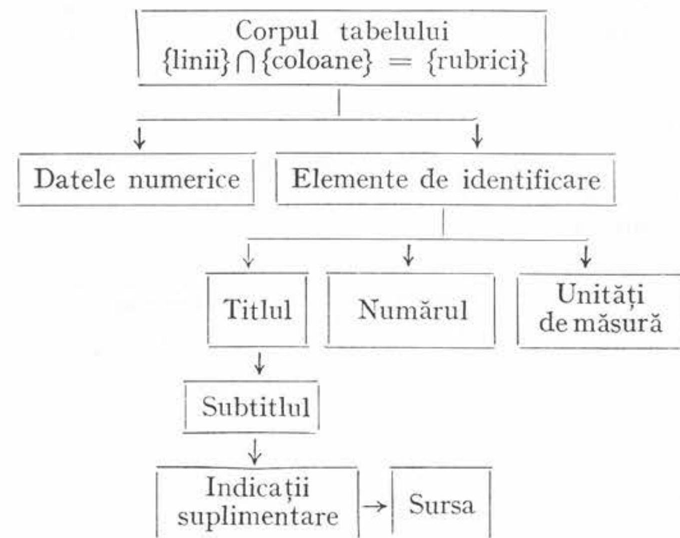
Să vedem acum felul în care se prezintă efectiv datele statistice deci, în definitiv, selecția cu care vom lucra mai departe.

Forma cea mai cunoscută și cea prin care unii dintre noi ne-am format o idee unilaterală despre statistică este *tabelul statistic*. El cuprinde populația luată în studiu și părțile ei componente, caracteristicile investigate cu valoarea lor numerică.

Alcătuirea unui tabel statistic „ca la carte“ nu este greu de făcut, dar pentru ca acesta să fie într-adevăr de folos, trebuie să ținem seama de următoarele elemente:

1°. Consultarea tabelului să nu implice studiul unor texte suplimentare, deoarece aceasta înseamnă că este incomplet și nu sugerează nici idei, nici concluzii;

2°. Să fie întocmit ca în exemplul de mai jos:



Iată un exemplu:

Țara	Producția	Consumul
	milioane tone	kg/cap de locuitor
S.U.A.	56,1	290
Japonia	16,0	212
Canada	13,1	205
R.F.G.	6,4	136
Finlanda	5,4	163

Tabelul 1: Producția și consumul de hîrtie și carton în unele țări ale lumii în anul 1973 (Sursa: Revista Economică, 34, 1975, p. 29).

Cînd întîlnim vreun tabel statistic, primul lucru pe care trebuie să-l facem este să-i citim titlul, să vedem despre ce este vorba. Titlul trebuie să conțină — într-o frază, de preferat, concisă — indicații privind obiectul tabelului, momentul de referință și — atunci cînd este comună — unitatea de măsură.

În cazul tabelului nostru, unitatea de măsură nu este comună și, în consecință, s-a procedat la întocmirea a două rubrici în care se specifică detaliile. Desigur, aceasta este la latitudinea cercetătorului, care mai avea și varianta:

Țara	Producția	Consumul
	(în milioane tone)	(în kg/cap locuitor)
⋮	⋮	⋮

Nici un mod nu este greșit: important este să se evite confuziile. Imediat după titlu ar urma subtitlul. Acesta, de obicei, apare atunci cînd titlul este prescurtat și sînt necesare unele lămuriri adiționale.

În continuare, în mod firesc, este necesar să se indice sursa, adică de unde, de către cine (eventual cum și în ce mod) au fost culese datele respective. Dacă cumva exemplul este „construit" în scopuri didactice, este de datoria autorului să menționeze aceasta prin calificativul „exemplu artificial" sau „date fictive" pentru a nu induce cititorii în eroare.

De exemplu, în manualul *Elemente de teoria probabilităților și statistică* de Gh. Mihoc și N. Micu, pentru cl. XII-a licee, (EDP, București, 1968), la pag. 90, autorii folosesc unele date „construite". Se indică însă imediat că datele respective sînt imaginare, pentru a-l lămuri pe cititor.

Desigur, și aceste „date fictive" trebuie să fie astfel alese încît să nu apară greutatea unei foi de hîrtie de scris, de exemplu, de ordinul kilogramelor, înălțimi de oameni de cinci metri etc., adică și aceste exemple „făcute" să nu se depărteze prea mult de realitate.

Indicarea sursei dă posibilitatea cititorului să verifice autenticitatea datelor numerice și să consulte și alte surse pentru comparare și documentare în plus.

Numărul tabelului este, evident, necesar pentru a identifica tabelul într-o carte, într-o colecție de tabele sau într-o publicație. Numărul tabelului este într-un fel „numele de familie unic" al tabelului respectiv. Astfel, în lucrarea noastră, dacă ne vom mai referi la tabelul 1, atunci îl vom detecta foarte ușor, răsfoind cartea pînă ce dăm de primul tabel.

Indicațiile suplimentare apar atunci cînd există unele situații speciale care nu rezultă din celelalte elemente ale tabelului.

Ele se scriu imediat după corpul tabelului sau după titlu, după preferință. Iată un exemplu:

Anul școlar / Țara	1972/1973	1973/1974
Bulgaria	3937	3763
Cehoslovacia	10834	10586
Cuba *	16123	16297
România	15495	15241
Ungaria	5505	5273

* fără școlile serale

Tabelul 2. Numărul unităților de învățămînt general de toate gradele în unele țări socialiste, la începutul anului școlar (Sursa: „Revista de Statistică", nr. 7, 1975, p. 58).

Anul	Oameni de știință și ingineri	Medici	Total
1966	3010	1465	4475
1967	6253	1993	8246
1968	5901	2041	7942
1969	6361	1923	8284
1970	9025	2211	11236
Total	30550	9633	40183

Tabelul 3. Intrări* de cadre științifico-tehnice și medicale din țările în curs de dezvoltare** în S.U.A. (Sursa: „Revista Economică", nr. 32, 1975, p. 26).

* persoane ce au primit oficial titlul de imigrant;
** țări din Asia, Africa și America Latină.

După cum se vede, pentru Cuba, cifrele indicînd unitățile respective nu includ școlile serale, fapt care, bineînțeles, nu aveam de unde să-l cunoaștem fără o indicație suplimentară.

Uneori, sînt necesare indicații suplimentare pentru a lămuri mai bine titlul. Acestea se pot trece imediat după titlu sau după indicarea sursei, după preferință. Iată un exemplu și pentru acest caz, în tabelul 3.

Tabelul mai conține, în plus, pe lîngă redarea în rubrici a datelor efective, și rubrici în care datele înscrise reprezintă rodul unor operații aritmetice cu datele respective. Regăsim în tabel și imaginea pe care mulți dintre noi o au despre tabelele statistice, unde, în final, însumarea pe linii trebuie să fie egală cu cea de pe coloane.

Aici, însumările se verifică, deoarece cercetătorul a dorit să obțină o imagine de ansamblu a situației analizate și, în plus, aici operațiile aritmetice se pot efectua fiindcă au sens. De pildă, în tabelul 1, însumarea datelor pe coloană nu înseamnă nici o informație utilă (ar fi că producția comună de hîrtie și carton a S.U.A., Japonia etc. este de „atîtea" milioane tone; dacă țările respective ar face parte dintr-o organizație economică comună, atunci o astfel de sumă ar prezenta un oarecare interes pentru comparația cu o altă organizație economică, la aceleași produse) în timp ce însumarea pe linii este, evident, o aberație.

Tabelul statistic trebuie „să vorbească" prin sine însuși. El trebuie să dea de gîndit asupra fenomenelor care se află la baza întocmirii lor, să stimuleze curiozitatea, spiritul de cercetare.

Iată, să ne uităm mai atent la tabelul 3. O privire superficială ne face să tragem doar concluzia că în decursul a cinci ani (1966 — 1970) un număr de 40 183 cadre calificate au emigrat, din țările în curs de dezvoltare, în Statele Unite ale Americii. La prima vedere, dacă ne menținem o gîndire de suprafață, cifrele

pot să nu ne impresioneze şi să nu ne spună mai nimic. Emigraţia şi imigraţia fac parte din fenomenele unor ţări încît mişcarea unor cadre calificate dintr-un colţ al planetei într-altul poate fi considerat de unii ca un caz particular în fenomenul emigraţie-imigraţie.

Totuşi, aceasta e doar la prima vedere: dincolo de cifre trebuie întotdeauna să intuim fenomenul specific.

Cifrele din tabelul 3 oglindesc statistic „transferul invers de competenţă" după însăşi expresia experţilor UNCTAD. Or, acest fenomen este unul dintre fenomenele contrare eforturilor de lichidare a subdezvoltării. Ce înseamnă 40 183 de cadre calificate pierdute de ţările „lumii a treia"? Evident, nu numai o simplă cifră, ci un întreg angrenaj de consecinţe negative pentru aceste ţări; să ne gîndim numai la pierderile băneşti legate de formarea unui specialist, de accentuarea decalajelor, căci exodul de „capital uman" spre alte ţări înseamnă, evident, întărirea, cîştigul de avantaje pentru acestea din urmă.

Pentru a căpăta o imagine concretă a acestor cifre să cităm cazul Indiei — una din ţările cu cel mai mare număr de specialişti de înaltă calificare ce au emigrat în S.U.A.

Conform unor studii efectuate de statisticieni şi economişti ai comisiilor specializate O.N.U., în 1970, pierderile pentru India se cifrau la aproximativ 17 000 dolari pentru un emigrant în domeniul ştiinţific şi învăţămîntului superior şi 32 000 dolari pentru un emigrant în domeniul tehnologic. Implicaţiile nu pot fi decît negative pentru ţările în curs de dezvoltare.

Un alt gen de tabel statistic este tabelul aşa-numit „comercial". Marile magazine, în scopul de a veni în ajutorul cumpărătorilor, întocmesc placarde în care sînt sintetizate caracteristici funcţionale, preţuri, modalităţi de cumpărare. Aceste tabele comerciale prezintă particularitatea că nu au număr şi dedesubtul sau deasupra lui nu scrie cuvîntul „tabel". Ele poartă doar un titlu, de obicei scris cu litere foarte mari.

În aranjarea noastră, o astfel de „placardă statistică" arată în felul următor:

Denumirea televizorului	Diagonala ecranului (cm)	Preţul (lei)	Aconto 15% (lei)	Rata lunară (24 rate) (lei)
Venus Olimp	47	2870	431	105
Modern	47	3050	458	110
Saturn Diana Opera Electra	59	3500	525	125
Clasic	59	3530	530	125
Opera 2 Diana 2	61	3550	533	126
Astronaut Lux	65	3960	594	140

Tabelul 4. Caracteristici ale unor televizoare de producţie internă şi înlesnirile lor de cumpărare. (Sursa: „Revista Economică", 21, 1975, p. 33).

Este cazul să dăm şi un exemplu de tabel statistic în care datele provin din domeniul tehnologic. De cele mai multe ori, aceste tabele privesc o singură caracteristică (un singur parametru) luată în considerare,

tabelul fiind denumit „simplu", ca în exemplul de mai jos:

Nr. burghiului (Ø = 8 mm; RW 180)	1	2	3	4	5	6	7	8	9
Durabilitatea (în minute)	21,56	22,65	10,12	22,08	26,20	18,43	24,24	25,12	21,31

Tabelul 5: Durabilitatea unor burghie elicoidale testate pe fontă F.C.18, duritate HB=170 ... 230. Avansul s=0,2 mm/rot, viteza de așchiere v = 28 m/min. (Sursa: Dorin, A., Vodă, V. Gh. *Eliminarea rezultatelor experimentale eronate pe baza unui criteriu statistic.* În: „Calitatea Producției și Metrologie", nr. 3, 1973, p. 158).

Din tabel, ne dăm seama că datele experimentale trecute în linia a doua sînt rodul unei încercări de durabilitate ale unor scule așchietoare, în particular al unor burghie elicoidale de diametru Ø = 8 mm, din oțel aliat tip RW 180. Încercarea — conform titlului tabelului — a fost efectuată pe un material tip fontă F.C.18 cm, duritatea HB = 170 ... 230, parametrii regimului de așchiere fiind respectiv: avansul = 0,2 mm/rot, iar viteza de așchiere 28 m/min.

Tabelul 5 conține toate elementele necesare unei informații rapide, încît consultarea unui text suplimentar devine necesară numai pentru un specialist în domeniul mașinilor-unelte sau prelucrării metalelor care dorește să aprofundeze problemele legate de durabilitatea sculelor.

Am afirmat că, în general, în studiul proceselor tehnologice apar tabele simple în care se sintetizează datele privind un singur parametru. Aceasta nu înseamnă că studiul proceselor industriale este simplu și că se ia în considerare numai un parametru. Se știe

că într-o întreprindere industrială, în diferitele procese tehnologice intervin tot felul de factori, adică, de fapt, există diverși parametri ai acestor procese: presiunea, temperatura, tensiunea curentului electric, conținutul diferitelor substanțe chimice, modul în care se execută diferitele operații mecanice (vezi chiar așchierea din tabel: tipul de material al burghiului, dimensiunea, materialul prelucrat etc.) ș.a.

În multe procese, acești parametri depind unul de altul sau se influențează, deci variații ale unuia implică în mod necesar variații ale celorlalți.

Deobicei însă parametrii respectivi se tabelează separat cu scopul concentrării atenției asupra fiecărui parametru în parte.

În cazurile concrete în care există dependențe de alți parametri sau în care caracteristica de interes este *multidimensională* (de pildă, la o piesă cilindrică ne interesează diametrul și lungimea sa) atunci, evident, tabelul va sintetiza în mod corespunzător datele experimentale respective.

Să trecem acum la un alt mod de prezentare a datelor statistice. Este vorba de *reprezentarea grafică*, cu care sîntem familiarizați din mijloacele de informare de masă: ziare, reviste, cărți, televiziune etc. Ceea ce oferă graficele este avantajul unei imagini sugestive a fenomenului luat în studiu. Ele reprezintă, într-un fel, „partea estetică" a statisticii descriptive. Aici, fantezia statisticianului în reprezentarea grafică a datelor statistice joacă un rol important.

Dar graficele nu se fac numai de frumusețe sau de dragul de a ne exersa talentul în desen. De multe ori, ele sînt chiar necesare, în special cînd dorim să ne formăm rapid o idee despre structura și tendințele unor fenomene sau procese.

Pentru cei dintre noi care nu au prea multe cunoștințe de statistică, graficele sînt de mare ajutor pentru că apelează în primul rînd la intuiție. Este ceva asemănător, poate, cu o problemă de geometrie plană: ea se

poate rezolva, dacă ne ambiționăm, fără să trasăm măcar o dreaptă. Dar, parcă altfel arată o astfel de problemă însoțită de o figură corect executată și folosind, eventual, culori. Criteriul estetic este aici secundar: o figură corectă, de cele mai multe ori sugerează însăși soluția problemei, fapt care nu este deloc de disprețuit căci, desigur, acesta este scopul.

N-o să părăsim subiectul fără a arăta cîteva modalități de reprezentare grafică a datelor statistice.

Una dintre modalitățile folosite este în acest sens *diagrama dreptunghiulară*: ea se caracterizează prin faptul că mărimea numerelor și a raporturilor dintre ele este exprimată cu ajutorul lungimii și lățimii unei suprafețe dreptunghiulare. Caracteristicile sau indicatorii luați în considerare se reprezintă prin dreptunghiuri ale căror arii sînt direct proporționale cu mărimile caracteristicilor respective.

În mod analog, se pot utiliza *diagramele pătrate* în care indicatorii respectivi sînt reprezentați prin pătrate ale căror arii sînt direct proporționale cu mărimea caracteristicilor considerate.

Cele mai folosite diagrame pentru compararea structurii unor colectivități sînt *diagramele circulare,* în care diferitele elemente ale colectivității apar ca sectoare de cerc.

Ca și la celelalte două diagrame, suprafața cercului este direct proporțională cu mărimea indicatorului respectiv deci ca să aflăm raza cercului, egalăm aria cercului (πR^2) cu mărimea caracteristicii studiate (fie ea C). Așadar, raza cercului este

$$R = \left[\frac{C}{\pi}\right]^{1/2}.$$

Diagramele circulare sînt foarte sugestive și ușor de construit. În unele cazuri, se pot utiliza mai multe diagrame circulare, în combinație, pentru descrierea unui fenomen.

Iată o astfel de combinație sugestivă:

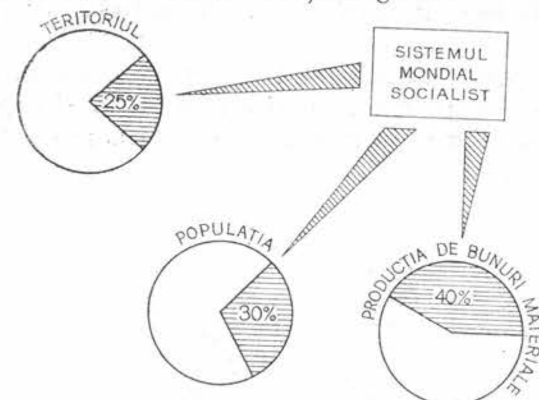


Fig. 3 — Ponderea sistemului mondial socialist (în procente) la trei indicatori fundamentali (Sursa: „Revista de Statistică", 5, 1975)

Acum să prezentăm și o diagramă circulară simplă:

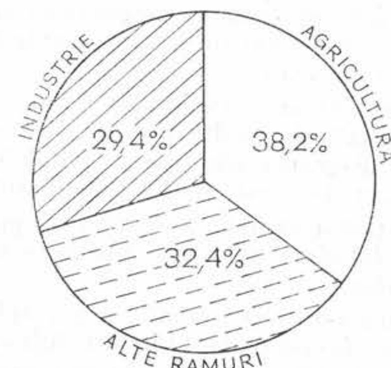


Fig. 4 — Structura populației ocupate a R. S. România (în procente) în anul 1975 (Sursa: Revista Economică, 33, 1975)

Figura 4 reprezintă o diagramă de structură; după cum se observă, ea este destul de simplă și suficient

de sugestivă, iar din această cauză o vom întîlni destul de des în lucrările statistice sau în materiale de informare.

Diagramele (indiferent dacă sînt circulare sau nu) se folosesc și pentru a înfățișa o prognoză și o previziune, nu numai pentru redarea unei situații existente.

De pildă:

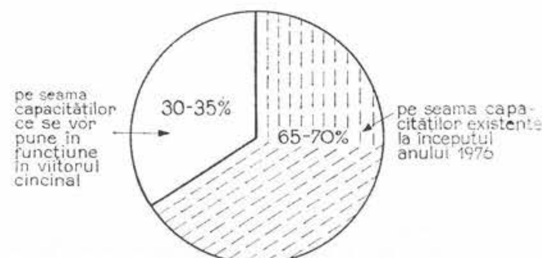


Fig. 5 — Realizarea producției industriale a anului 1980 (Sursa: Anexa la „Revista de Statistică", 10, 1974)

O să încheiem prezentarea diagramelor statistice cu un tip de diagramă oarecum familiar tuturor și anume cea *prin coloane paralele.*

Fiindcă petrolul este în zilele noastre una din materiile prime „la modă" din multe puncte de vedere, iată, în figura 6, o diagramă prin coloane paralele înfățișînd un aspect legat de acest produs „fierbinte".

Caracteristica esențială a diagramei prin coloane paralele este folosirea unei scări numerice la reprezentarea coloanelor.

Astfel, dacă s-au reprezentat, de exemplu, cele 100 de *lire sterline* investite în Orientul Mijlociu printr-o coloană înaltă de 2 mm, atunci coloana corespunzătoare investiției din Nigeria (£. 400) trebuie să fie înaltă de 8 mm etc., astfel încît să obținem o imagine fidelă a situației reale, pentru punerea în valoare a discrepanțelor etc.

Și diagrama, nu numai tabelul trebuie să ne facă să reflectăm la fondul problemelor.

Diagrama are în plus avantajul „pictural" care ne poate sugera intuitiv idei sau chiar concluzii.

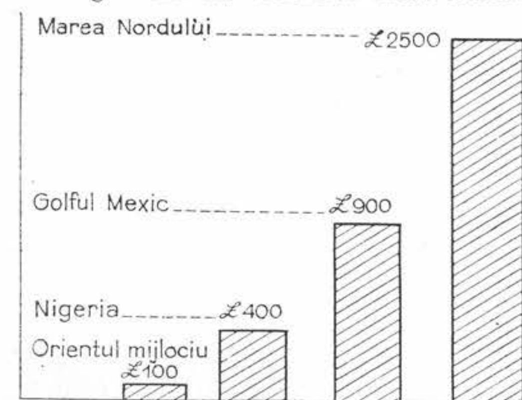
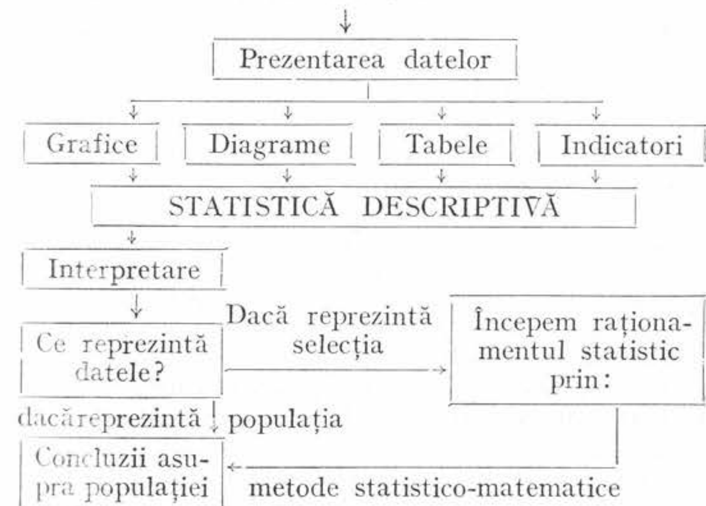


Fig. 6 — Investiții ocupate pe 1 baril de petrol producție zilnică, în lire sterline, în diferite zone ale lumii, în anul 1975 (Sursa: Revista Economică, 33, 1975)

Sîntem acum în măsură să completăm schema noastră cu următoarele elemente:



1.4. FALSE PĂRERI ASUPRA STATISTICII

„... nu trebuie să uităm că există o necesitate imperioasă de a îmbunătăți înțelegerea și aprecierea statisticii de către marele public".

B. H. Colvin [1]

Ca în multe alte domenii, părerile false asupra unei situații, asupra unei probleme, asupra unei ramuri de activitate umană etc. pot proveni din cauza mai multor factori.

În primul rînd, un factor universal este *necunoașterea problemei*. Este foarte clar faptul că necunoscînd un anumit subiect, dar totuși încercînd să emitem judecăți despre el, acestea vor avea o șansă foarte mică să corespundă realității.

Nu trebuie aici să ne gîndim neapărat la marele public căruia nu i se poate, în mod normal, pretinde să fie expert în statistică matematică. Trebuie să-i avem în vedere pe specialiștii din alte domenii precum și chiar pe cei din domeniul matematicii.

Pe de altă parte, o informare generală asupra trăsăturilor esențiale ale unui anumit domeniu, asupra frumuseții, greutăților, obstacolelor ce se pot întîlni în acel domeniu, asupra importanței reale și a utilității domeniului respectiv în contextul dezvoltării economico-sociale a țării, este o datorie.

Imaginea despre care vorbeam la început — aceea a adunării unor numere pe linii sau coloane —, ca trăsătură fundamentală a muncii statisticianului, persistă nu numai la „profan", care, după cum repetăm, nu este de fel vinovat de o asemenea părere, ci și în cercuri de specialiști din alte domenii.

Este interesant de urmărit și sensul care se acordă denumirii de „statistician". În mod uzual, statistician

[1] Colvin, B. H., *A wiew from industry*, Proc. First Canad. Conf. in Applied Statistics, „Statistics '71 Canada", 1971, p. 59.

este considerat acel individ care se ocupă cu calcule (care în concepția generală nu depășesc operațiile aritmetice), scopul său final fiind obținerea aceleiași sume pe liniile și coloanele unui tabel!

Pînă nu demult, în nomenclatorul profesiunilor din multe țări, aceasta se înțelegea prin statistician. Azi, statisticianul este persoana cu studii de specialitate, ca și ciberneticianul, analistul de sisteme, informaticianul etc.

Statistica matematică constituie baza teoretică a profesiunii de statistician. Experiența practică, pe de altă parte, constituie condiția esențială a unui bun statistician, cu adevărat util societății.

După cum spunea W. E. Deming — unul din clasicii statisticii matematice, cunoscut printre altele și prin lucrările sale de filozofie a statisticii — „statisticianul are nevoie de anumite principii practice pentru aplicarea efectivă a științei statisticii. Cunoașterea teoriei statistice este necesară dar nu și suficientă. Teoria statistică nu conține în ea însăși și modelul efectiv de aplicare a ei"[1].

În legătură cu aceasta, ceea ce a constituit un alt factor ce a contribuit la o falsă înțelegere a statisticii a fost nesesizarea faptului că în epoca modernă, știința și tehnologia au o natură interdisciplinară. Acest proces de întrepătrundere s-a accelerat îndeosebi după cel de-al doilea război mondial, cînd multe cunoștințe științifice și tehnice care s-au dovedit utile în timpul războiului au fost imediat adaptate în activitatea comercială și industrie.

Au apărut astfel cibernetica, bionica, cercetarea operațională, știința comportamentului grupelor de indivizi (behavioral science).

Clasificarea clasică a științei și tehnologiei a căzut treptat. Noua dezvoltare a produs un nou concept al științei ca un arbore cu numeroase ramificații — cunoașterea și aplicațiile ei.

[1] Deming, W. E., *Principles of profesional statistical practice*, „Ann. Math. Statist.", vol. 36, nr. 6, 1966, p. 1883.

„Dacă nu recunoaştem acest nou concept — spunea recent cercetătorul japonez Tareshi Suzuki [1] — atunci nu ne putem aştepta la prea mult progres".

Este foarte interesant să subliniem faptul că încă în secolul trecut, agronomul român Ion Ionescu de la Brad — printre alţii, unul din pionierii învăţămîntului statistic în ţara noastră — scria:

„Statistica pur matematică cuprinde, pe de o parte, calculul probabilităţilor ce aduce aminte de Laplace, Lacroix, Poisson, pe de altă parte, aceea pe care Quetelet a numit-o statistica morală (statistica socială — n.n.).

Statistica morală este o aplicare a acestor metode şi a altor mijloace matematice la previziunea fenomenelor specific umane ...

Acei ce sînt numai matematicieni, din lipsă de cunoştinţe economice ... nu au dat lucrurilor toată însemnătatea socială ce le trebuia acordată" (s.n). [2]

Iată deci că un român şi încă în secolul trecut şi-a dat foarte bine seama de caracterul interdisciplinar al statisticii. Desigur, el privea numai o ramură a ştiinţei statistice, dar semnificativ este faptul că el a sesizat magistral pericolul la care sînt supuşi matematicienii care se menţin în sfera teoriei pure, refuzînd contactul cu problemele lumii înconjurătoare.

Acest pericol este astăzi dublu: pe de o parte, pentru înşişi matematicienii în cauză, pe de altă parte, pentru deserviciul pe care aceştia îl pot aduce prin desconsiderarea ramurilor mai „pămînteşti" sau prin desconsiderarea părţii aplicative a propriei ramuri de matematică.

Ne gîndim aici la matematicianul care a ales drept profesiune statistica. Pericolul pentru soarta statisticii

---

[1] Suzuki, T. *Cybernetic relationship to quality control,* Proc. EOQC — IAQ Conference—Venice 16/17-th sept. 1975, vol. I, 1975, p. 165.

[2] Textul este adaptat în limba română modernă. Ion Ionescu de la Brad scria, de exemplu, „curatu" în loc de „pur", „soiu" în loc de „specific" sau de „fel" (*Studii de statistică,* Ed. D.C.S., Bucureşti, 1969, p. 13).

stă în atitudinea lui faţă de înţelegerea naturii esenţiale a propriei lui ştiinţe.

Poate aceste probleme nu par importante la prima vedere însă în realitate ele sînt importante, îndeosebi în ceea ce priveşte tînăra generaţie. Orientarea ei corectă, alături de orientarea marelui public este mai importantă la un moment dat decît descoperirea unei teoreme.

Iată ce remarca un statistician canadian la primul congres de statistică aplicată „Statistics 1971, Canada":

„În eforturile noastre de a îmbunătăţi educaţia ştiinţifică a statisticienilor nu trebuie să uităm că există o necesitate imperioasă de a îmbunătăţi înţelegerea şi aprecierea statisticii de către marele public ...

Mulţi dintre noi au observat eforturile mari care s-au depus în această direcţie în ultimele două decenii. Este clar acum că aceste eforturi deşi substanţiale au fost oarecum ineficiente.

Cum să desfăşurăm acţiunea în viitor, mi se pare că aceasta depinde foarte mult de *modul în care vom şti să educăm generaţia viitoare de statisticieni (s.n.)*

Dacă ei vor fi interesaţi şi capabili să scrie articole de popularizare care să lămurească pe omul obişnuit, pe omul dintr-o funcţie de conducere, atunci putem spera într-un mai mare progres"

(Colvin, B. H., *A view from industry,* op. cit., p. 59.)

Pentru aceasta este, bineînţeles, necesar ca însuşi cel ce lucrează în domeniul statisticii să aibă o înţelegere foarte clară a esenţei acestui domeniu.

Or, „ ... natura esenţială a statisticii este aceea de slujitor al ştiinţei. Cea mai importantă sarcină constă în cuantificarea incertitudinii — fapt inerent în orice ştiinţă experimentală. A doua sarcină importantă a statisticii, ca slujitor al ştiinţei constă în acumularea unor „provizii" de modele şi tehnici de analiză a datelor care să ajute pe cercetătorii din alte domenii să-şi poată rezolva problemele" [1].

---

[1] Rutherford, J. R., *Aims for training programs in statistics,* Proc. First. Canadian conference in Applied Statistics „Statistics 71, Canada", 1971, p. 327.

Aici trebuie iar să punem în evidență un factor generator de păreri false asupra statisticii și anume *supraestimarea* ei. Această supraestimare a alimentat fantezia umoriștilor care l-au imaginat pe statistician drept „omul care poate demonstra orice din nimic".

Este binecunoscut (cel puțin printre statisticieni!) umoristul canadian S. Leacock care scria: „În timpurile vechi oamenii nu aveau statistici, așa că au trebuit să recurgă la minciuni. De aici marile exagerări ale literaturii primitive — giganți, miracole sau minciuni! Ei făceau așa ceva cu minciuni, iar noi facem așa ceva cu statistică; dar e cam același lucru".

Nici această părere nu a apărut întîmplător: ea este rodul exagerărilor asupra posibilităților și rolului statisticii în celelalte domenii de activitate umană.

Succesele înregistrate de statistica matematică prin aplicarea metodelor sale într-un evantai foarte larg de domenii a condus la apariția unei „euforii" a statisticii. S-a crezut la un moment dat că statistica poate substitui metodele specifice domeniului în care aceasta este aplicată. Dacă vom aplica metode statistice în biologie și medicină, în nici un caz nu vom vindeca pacienții cu ajutorul statisticii. Ceea ce poate face statistica aici este, de pildă, compararea eficacității diferitelor vaccinuri, determinarea corelațiilor între un agent presupus patogen și apariția unor anumite simptome de boală etc.

Toate acestea sînt de un real folos, dar ca să devină efective trebuie, în mod obligatoriu, combinate cu cunoașterea specifică biologiei sau medicinii.

Este binecunoscut faptul că statistica și calculatoarele își aduc contribuția la bătălia împotriva cancerului. Dar nici statistica și nici calculatoarele nu ne vor vindeca de acest flagel.

La fel stau lucrurile și în alte domenii: în industrie, statistica nu va produce niciodată. Ea devine însă o forță de producție doar atunci cînd este folosită pentru a ne ajuta să producem mai bine.

Din această cauză, statisticianul nu-l va putea înlocui niciodată pe inginer, pe maistru sau pe muncitor. El trebuie însă să se alăture acestei grupe și împreună să conlucreze la diagnosticarea proceselor de producție „bolnave".

O analiză amănunțită a unor probleme discutate de noi aici a făcut-o profesorul George Theiler de la Universitatea din București într-un ciclu de articole început în decembrie 1974 și publicate în „Revista de Statistică" — organ al Direcției Centrale de Statistică și al Societății de științe economice din R.S. România. Ciclul se intitulează „Cu privire la posibilitățile de folosire eficientă a metodelor statisticii matematice în studiul fenomenelor economice și sociale" (primul articol a apărut în „Revista de Statistică", nr. 12, 1974, pp. 7—16). Recomandăm, în special, primul dintre aceste articole. Nivelul cunoștințelor de matematică nu depășește cl. XII de liceu. Importante sînt însă vorbele. Într-o lucrare de statistică acestea nu trebuie niciodată neglijate. Afirmații ca: „n-are formule deci nu e bun" oglindesc neînțelegerea totală a unei discipline care, ca și fizica sau chimia nu poate fi desprinsă de fenomenele reale ce le modelează sau le descrie. Pentru exemplificare, vom extrage din el un scurt pasaj în legătură directă cu ceea ce dezbatem aici: false păreri asupra statisticii.

Iată cum sună acel pasaj:

„Privitor la statistică, va trebui ca lucrurile să fie bine precizate spre a nu se ajunge la situații dintre acelea care au dus la definirea statisticii ca o disciplină cu ajutorul căreia se poate dovedi orice.

Aici nu este vorba de o glumă, ci de rezultatele unor procedee greșite, avînd la bază date eterogene, concluzii forțate dincolo de implicațiile datelor de observație.

Să ne închipuim astfel că urmărim activitatea din două licee. Primul, cu un corp profesoral bine pregătit și cu elevi în majoritate corespunzători, în care materiile programate se predau în totalitatea lor, chiar cu unele completări și în care notele se pun pe baza cunoștințelor reale.

Un al doilea, în care corpul profesoral este mai puțin pregătit, elevii mai slabi, în care se trece cu ușurință peste anumite părți mai

dificile din materii, iar notele se pun cu indulgență, în genere nereprezentînd cunoștințele reale ale elevilor. Să presupunem acum că anumite foruri superioare se orientează după proporția de promovați și media globală de promovare. Urmarea? Liceul mai puțin corespunzător va apărea mai bun, iar liceul cu adevărat bun — mai slab. Iată la ce imagine deformată a realității se poate ajunge folosind date eterogene și indicatori neconcludenți...

În locul a două licee ne-am fi putut imagina două întreprinderi cu același specific. Ne-am fi putut pune problema modului în care se depășește planul, făcîndu-se abstracție de calitate.

Nu s-ar fi putut ajunge și aici la aceeași situație paradoxală? *Se vede deci că nu cu ajutorul statisticii se poate demonstra orice, ci cu ajutorul unor procedee greșite sau interpretări eronate în care statistica este luată drept paravan. În astfel de cazuri nu este vorba de statistică, ci de cu totul altceva*". (s.n.).

Într-adevăr, este vorba de cu totul altceva, care se numește incompetență.

Căci tot același autor ne avertizează la început:

„Statistica este o disciplină a cărei aplicare utilă cere competență: o justă înțelegere a conceptului de statistică, a sferei de aplicabilitate a posibilităților și limitelor sale".

Granițele metodelor statistice — ale inferenței statistice — au fost pe larg analizate de către W. E. Deming într-un articol intitulat chiar astfel: „Granițele inferenței statistice"[1] care poate fi considerată ca o lucrare de filozofie a statisticii. Astfel de lucrări nu se povestesc; ele trebuie citite. Și nu numai de cei care vor să se apuce de statistică.

Nu vom cita decît o frază din articolul lui Deming:

„Alături de limitările teoriei statistice, statisticianul nu trebuie desigur să uite puterea ei".

Cîte ceva despre această putere vom găsi în paragraful ce urmează.

---

[1] Deming, W. E., *Boundaries of statistical inference*, În: „New Developments in Survey Sampling", N. L. Johnson and H. Smith Jr. Editors, Wiley-Interscience, New York, 1969, p. 652.

1.5. NECESITATEA FOLOSIRII STATISTICII

„Ca domenii de aplicare, considerăm că nu există ramură a economiei naționale în care să nu fie necesară folosirea metodelor statistico-matematice".

Manea Mănescu[1]

Dacă am încheiat paragraful precedent cu un citat, atunci pentru a menține tonul îl vom începe pe acesta tot cu un citat.

De data aceasta nu din Deming, ci dintr-un alt clasic în viață al statisticii și anume profesorul Irving Burr[2]:

„Ori de cîte ori rezultatele unui experiment variază într-un mod neprevăzut sau într-o manieră întîmplătoare, avem de-a face cu date statistice *fie că ne place sau nu, fie că știm aceasta sau nu* (s.n.).

Dacă așa stau lucrurile, atunci este necesar să folosim cele mai potrivite metode statistice pentru a le analiza, astfel încît să fim obiectivi și să evităm capcanele și interpretările greșite".

Ideea nu este, bineînțeles, nouă; ea este exprimată însă foarte plastic și oglindește un aspect pe care, într-un fel sau altul, l-am mai întîlnit și în cele discutate anterior, anume că folosirea statisticii este o necesitate obiectivă — fie că ne place sau nu — cum spune profesorul Burr[2].

Este interesant de amintit aici că marii clasici ai materialismului dialectic și istoric au acordat statisticii — chiar așa insuficient dezvoltată la vremea

---

[1] *Introducere* la: Rancu, N., Tövissi, L., *Analiza statistico-matematică a calității producției industriale*, Edit. Științifică, București, 1964, p. 14.

[2] Burr, I. W., *Applied Statistical Method*, Academic Press, New York, 1974, p. 13.

aceea — importanța cuvenită, intuind posibilitățile acesteia de a contribui la lupta clasei muncitoare [1].

Astfel, Marx scria, în 1877:

„... una dintre măsurile importante propuse de noi pentru unitatea internațională a acțiunilor este cercetarea statistică a situației clasei muncitoare din toate țările, care trebuie făcută de însăși clasa muncitoare".

(K. Marx, Fr. Engels, *Despre Sindicate*, Ed. C.G.M., București, 1947, p. 146).

Desigur, Marx nu se referea la statistica matematică, ci la cea social-economică, deoarece în acea perioadă, după cum am mai subliniat, metodele matematice specifice statisticii abia se nășteau în timp ce statistica social-economică, deși mai mult cu caracter descriptiv, avea un trecut destul de vechi și își dovedise utilitatea.

Mai tîrziu, V.I. Lenin a apreciat statistica social-economică drept „unul dintre cele mai puternice instrumente de cunoaștere socială"·

(Lenin, V. I., *Opere*, vol. 16, ESPLP, București, 1957, p. 435).

Evoluția istorică, care a adus cu sine dezvoltarea metodelor matematice și implicit a statisticii matematice, nu a făcut decît să confirme părerea lui Lenin.

Este bine să arătăm aici că și la noi în țară au existat gînditori luminați care și-au dat seama de necesitatea utilizării statisticii în domeniul social-economic.

Astfel, Gheorghe D. Creangă (1874—1940) economist și statistician care și-a consacrat întreaga viață studierii problemelor legate de agricultura țării noastre, scria în 1907:

„Statistica trebuie să-și întindă activitatea mai cu seamă — și mai ales în țările în care acum se dezvoltă — în direcțiunile în care se manifestă un rău de ordin general, de natură a periclita progresul țării, *lăsînd pe linia a doua cercetările mai puțin importante și care*

[1] Mănescu, G., *Statistica și mișcarea muncitorească*, In: „Revista de Statistică", nr. 4, 1971, pp. 3—16.

unele sînt mai mult de natură a satisface o curiozitate decît a remedia un rău general" [1]. (s.n.)

Continuînd în același spirit, el adaugă:

„Statistica trebuie să fie deci mai mult practică. Astăzi cifrele sînt cel mai bun sfătuitor al omului inteligent. Ele descoperă cauzele și ne arată leacul bolilor economice și sociale ...

Fără statistică — știința cifrelor — nu știm ce am fost, ce sîntem și ce vom fi".

O bună parte din argumentele care motivează necesitatea folosirii statisticii într-unul sau altul dintre domeniile cunoașterii umane le-am expus în introducere și pe specific le veți găsi detaliate în capitolele ce urmează.

Vom încheia această serie de „colaje" în sprijinul ideii noastre, cu o observație făcută cu aproape două decenii în urmă de către profesorul Manea Mănescu:

„Fără un vast material documentar cu privire la istoria economiei naționale și a statisticii autohtone este greu de elaborat o istorie a poporului" („Revista de Statistică", nr. 1, 1958, pag. 14).

Importanța și necesitatea statisticii se dovedește deci a fi nu numai pur matematică. Ea afectează toate sferele activității noastre. Este un univers din a cărui structură facem parte și prin urmare nu-l putem ignora.

[1] Creangă, Gh. D., *Învoielile agricole în România în vigoare pe anul 1906*, București, 1907 (citat din V. Bulgaru „Terra Nostra", 1973, p. 459).

Capitolul II

STATISTICA, INSTRUMENT PRACTIC DE ACȚIUNE

„... metodele statistice trebuie să poată fi aplicabile pentru analiza datelor fără vreo cunoaștere specială a statisticii".

Takeshi Suzuki [1]

Dacă urmărim atent dezvoltarea istorică a statisticii, pe lîngă alte lucruri, vom constata și unul care ne atrage atenția și anume că statistica a primit de-a lungul timpului diferite denumiri. La început, nedesprinsă de teoria probabilităților, era inclusă în această denumire.

Ceva mai recent s-a numit „teoria sondajelor", apoi pur și simplu „statistică".

Și mai recent, entuziaști ai noilor discipline ca informatica, teoria sistemelor, cibernetica sau cercetarea operațională au căutat pentru statistică — sau pentru părți ale ei — una dintre aceste denumiri.

Ceea ce a remarcat Deming este însă un fapt esențial: noile nume pentru o disciplină sau o profesie nu descoperă nici probleme și nici soluții. Denumirea este lucrul cel mai puțin important. Ceea ce este cu adevărat important este acțiunea, sub orice nume s-ar desfășura aceasta.

Vom căuta în acest capitol să vă înfățișăm modul în care statistica a devenit și este un instrument practic de acțiune.

Respectînd, din punct de vedere istoric, etapele de dezvoltare a statisticii vom analiza mai întîi aspecte ale aplicațiilor sale în domeniul vieții sociale iar apoi legăturile ei cu industria, cu fizica și cu chimia.

---

[1] Idem op. cit., p. 155.

2.1. STATISTICA ȘI DOMENIUL VIEȚII SOCIALE

A. Primele aplicații ale statisticii

„Statistica poate fi un ajutor, dar niciodată nu va substitui o gîndire corectă și clară. Din punctul de vedere al cercetătorului în științele sociale, ea este numai un instrument".

Blalock, H. M.[1]

Spuneam mai înainte că statistica a apărut în legătură cu fenomene ale vieții sociale și că unii termeni științifici din statistica de azi s-au păstrat din perioadele în care obiectul de bază al statisticii îl formau problemele legate de populația umană, de demografie sub multiplele ei aspecte.

Înregistrări ale diferitelor fenomene ca nașteri, dispariții prin moarte, lucrători într-o anumită branșă, situații privind comerțul, armata, finanțele se făceau din timpuri destul de vechi. Ceea ce nu se făcea însă era interpretarea acestor date, pentru că lipsea instrumentul matematic corespunzător.

De asemenea, acuratețea datelor nu era întotdeauna „la înălțime". Unele înregistrări erau incomplete și necentralizate, ceea ce făcea ca imaginea de ansamblu asupra unui fenomen să fie foarte vagă mai ales că singura interpretare ce se putea da în acele vremuri era strict legată de compararea ca atare a unor numere.

Din cîte se cunoaște, primul care a încercat o interpretare a unor fenomene sociale și biologice de masă pe baza unor date statistice a fost un mic negustor englez pe nume John Graunt (1620—1674).

Faptele s-au petrecut astfel: administrația orașului Londra a început să țină săptămînal evidența naște-

---

[1] Blalock, H. M., *Social Statistics*, Mc. Graw-Hill Book Co., New York, 1960, p. 13.

rilor şi a înmormîntărilor, în special datorită unei mari epidemii de ciumă care dereglase echilibrul populaţiei oraşului în acel timp.

Graunt a studiat listele privind mortalitatea în Londra, liste ce pe vremea aceea erau păstrate în diversele parohii ale oraşului.

În 1662 apare lucrarea sa intitulată *Observaţii naturale şi politice conţinute în următorul index şi efectuate pe baza listelor de mortalitate.* Titlul este cam lung, poate, dar aceasta este prima lucrare de demografie, prima lucrare în care statistica apare nu numai ca o simplă colecţie de tabele, ci şi ca un instrument de analiză şi prognoză în acelaşi timp.

Să menţionăm că Graunt nu era matematician; în această direcţie el poate fi considerat un amator. Ceea ce avea însă Graunt era un deosebit simţ al observaţiei şi al realităţii.

Iată unele din observaţiile sale privind fenomenele demografice ale perioadei respective în Londra:

a°. numărul născuţilor de sex masculin este mai mare decît al născuţilor de sex feminin, dar ca proporţie nu este mai mare decît cu 1/13 din născuţii de sex feminin;

b°. datorită faptului că numărul bărbaţilor adulţi este aproximativ egal cu cel al femeilor adulte, monogamia pare mai în concordanţă cu legile naturii;

c°. numărul cazurilor de moarte violentă este sensibil mai mare la bărbaţi decît la femei.

Un alt lucru interesant pe care Graunt caută să-l rezolve este problema estimării populaţiei Londrei.

Trebuie amintit că pe vremea aceea nu se ţinea o evidenţă a populaţiei ca în timpurile noastre. „Debandada" — cuvîntul nu este deloc tare — caracteriza fenomenul mişcării populaţiei.

Graunt a încercat să folosească pentru prima oară logica statistică pentru a interpreta datele existente.

Judecata sa se baza, în esenţă, pe următoarele fapte:

— numărul de naşteri era cunoscut din listele parohiilor;

— se cunoştea aproximativ fertilitatea femeilor (exprimată mai mult prin vîrsta fertilă).

Din această informaţie se putea *extrapola* (termenul e pretenţios oarecum, dar exprimă mai bine situaţia) numărul femeilor capabile de naştere.

Pe baza acestei informaţii se poate face o idee despre numărul aproximativ de familii ce se puteau forma şi totodată se putea da o estimaţie în medie a membrilor componenţi ai unei familii.

În fine, cu ajutorul acestei situaţii se putea trece la estimarea întregii populaţii.

Deşi destul de primitivă şi bazată probabil pe date ce nu meritau întotdeauna cea mai deplină încredere, metoda lui Graunt poate fi considerată prima metodă statistică aplicată cunoscută.

Una din problemele interesante de care s-a ocupat Graunt este aceea de a vedea cîte persoane dintr-un lot de 100 mor într-un număr dat de decade.

Vîrsta (ani)	Supravieţuitori (nr. indivizi)
0	100
6	64
16	40
26	25
36	16
46	10
56	6
66	3
76	1

Tabelul 6: Numărul de supravieţuitori din 100 de indivizi in diferite decade (Londra, sec. XVII). (Sursa: Hacking, I. *The Emergence of Probability*, Cambridge Univ. Press, 1975, p. 108).

El dă și un tabel care a suscitat ulterior numeroase comentarii cu privire la modul în care a fost alcătuit (vezi tabelul 6).

Hacking (1975) arată că, în fond, tabelul lui Graunt se deduce ușor prin rezolvarea ecuației:

$$64\,(1-p)^7 = 1, \qquad (5)$$

adică, presupunînd că 64 de indivizi supraviețuiesc la etatea de 6 ani și numai unul la 76 de ani, p reprezentînd proporția de indivizi ce dispar în decursul unei decade, proporție presupusă constantă în fiecare decadă.

Cum apare ecuația, este simplu de explicat: astfel, 7 reprezintă numărul de decade 6—16, 16—26, ..., 66—76; $64\,(1-p)$ reprezintă numărul de supraviețuitori în prima decadă.

Atunci numărul celor ce supraviețuiesc în a doua decadă este:

$$64\,(1-p) - 64\,p(1-p) = 64\,(1-p)^2, \qquad (6)$$

adică din numărul de supraviețuitori din prima decadă se scade numărul celor ce dispar în a doua decadă.

După șapte decade se obține $64\,(1-p)^7$ și cum s-a presupus că o singură persoană atinge maximum 76 de ani atunci rezultă ecuația (5).

Logaritmînd avem:

$$\lg 64 + 7 \lg (1-p) = 0 \qquad (7)$$

sau:

$$\lg (1-p) = -\frac{1}{7} \lg 64 = -0{,}2580.$$

Sau, mai departe:

$$(1-p) = 0{,}625$$

adică:

$$p = 0{,}375 = \frac{3}{8}. \qquad (9)$$

Acum nu are prea mare importanță dacă datele lui Graunt erau foarte exacte sau nu. Ținînd cont de anumiți factori economico-sociali, se pare că el nu era prea departe de situația reală.

Mortalitatea infantilă era, în acele vremuri, într-adevăr, foarte mare iar media de viață foarte scăzută; desigur, în zilele noastre asemenea cifre par ciudate, dar să nu uităm nici un moment timpul în care ele sînt valabile.

Iată deci care au fost începuturile statisticii aplicate, într-o perioadă în care teoria probabilităților abia se forma ca știință, și cînd însăși statistica nu prea știa bine care-i erau scopurile, metodele și posibilitățile. De fapt, pe acea vreme, statistica purta numele — azi bizar — de „aritmetică politică" (political arithmetic).

Societatea feudală, fărămițată sub toate aspectele, nu putea aprecia la adevărata lor valoare cunoștințele demografice. Preocupările științifice rămîneau la latitudinea unor minți luminate, care căutau să ducă mai departe cunoașterea legilor naturii și ale societății. Pe de altă parte, dezvoltarea societății a impus acordarea unei atenții mai mari problemelor demografice, colectării datelor privind populația sub diferite aspecte și analizării lor.

În această perioadă, industria, ca factor stimulator al științei în general se menține încă în umbră, datorită nivelului scăzut la care se afla.

În 1669 se înființează prima companie de asigurări și astfel încep să apară probleme legate de durata medie de viață, de extreme (durata maximă de viață), de rentele anuale ce trebuie plătite etc.

Începe să se pună baza așa-numitei „matematici actuariale", care, în fond, nu reprezintă altceva decît teoria probabilităților și statistică matematică aplicate în domeniul asigurărilor, cu toate aspectele legate de acestea.

Edmond Halley, celebrul astronom care a descoperit cometa ce-i poartă azi numele, a publicat în 1693 un articol privind rata mortalității. La sugestia lui Leibniz, el a folosit „curioase tabele de nașteri și decese" ale orașului Breslau (azi Wroclaw, R.P. Polonă). Ceea ce a mai încercat Halley a fost — pe baza acestor tabele de mortalitate — și stabilirea prețurilor pentru rentele anuale la asigurările de viață.

Teoria statistică a mortalității se dezvoltă în paralel cu teoria probabilităților.

Ceea ce este important să subliniem este faptul că această teorie a generat mai tîrziu instrumentul matematic pentru *teoria siguranței în funcționare a sistemelor tehnice* sau teoria „fiabilității" (fr. *fiabilité* = siguranță).

Iată, în varianta modernă, cum se pune problema determinării ratei mortalității cu care se ocupau predecesorii epocii moderne.

Să presupunem că luăm în studiu o colectivitate de indivizi de aceeași vîrstă x (considerată în ani, bineînțeles) și fie $P(x, y)$ probabilitatea ca unul dintre acești oameni luat la întîmplare, să fie încă în viață la împlinirea vîrstei y (evident, $y > x$).

Nu o să mai insistăm aici ce este probabilitatea unui eveniment și care-i sînt proprietățile.

Revenind deci, să notăm prin $Q(x, y)$ probabilitatea contrară, adică probabilitatea ce corespunde evenimentului:

{omul moare înainte de a împlini vîrsta y}

În mod natural:

$$P(x, y) + Q(x, y) = 1, \qquad (10)$$

deci am făcut uz de proprietatea probabilităților atașate evenimentelor contrare.

Să considerăm acum intervalul de timp $(x, x + \mathrm{d}x)$.

Intuiția ne spune că probabilitatea ca individul considerat de vîrsta x (viu la împlinirea vîrstei x) să moară la o vîrstă cuprinsă între x și $x + \mathrm{d}x$ ani, este foarte mică.

Aici „dx" are semnificația de variație mică a lui x, semnificație pe care ați întîlnit-o în problemele de analiză matematică sau fizică, astfel încît ipoteza ca probabilitatea respectivă să fie foarte mică este rezonabilă.

Că nu este nulă, se poate ușor argumenta printr-un contraexemplu: este cunoscut cazul mai multor indivizi care decedează pe stadioane, la meciurile de fotbal, din cauza unui stop cardiac provocat de o bucurie sau o supărare prea puternică, generată la rîndul ei de înscrierea unui gol.

Ei bine, consultînd listele cu astfel de decese, se poate găsi un caz care chiar cu o zi înaintea meciului fatal își serbase ziua de naștere.

Atunci, pare natural ca funcția $\lambda(x)$ definită astfel:

$$\lambda(x) = \frac{Q(x, x + \mathrm{d}x)}{\mathrm{d}x} \qquad (11)$$

să o numim *intensitate* sau *rată a mortalității.*

Dacă ne rememorăm ce înseamnă $Q(x, x + \mathrm{d}x)$, atunci vom găsi această definiție realistă.

Acum să apelăm la un minim de cunoștințe de analiză matematică. Cum va trebui să presupunem că este funcția $\lambda(x)$?

Dacă ne gîndim puțin, va trebui să o presupunem continuă și *crescătoare* fiindcă probabilitatea de moarte crește, evident, odată cu înaintarea în vîrstă.

Același lucru se întîmplă și cu „mortalitatea obiectelor" de care ne folosim zilnic. Dacă televizorul la care ne uităm are o vechime de 10 ani, atunci probabilitatea ca el să se defecteze „pe vecie" este mai mare decît aceeași probabilitate pentru un televizor cumpărat numai de un an.

Cineva poate replica astfel: cunosc unele cazuri cînd lucruri noi, abia cumpărate s-au defectat foarte repede.

Aici se întîmplă ceva similar cu fenomenul de „mortalitate infantilă". El se întîlnește mai des la componentele electronice și este numit *proces de defectare timpurie.* Din această cauză sînt prevăzute și termenele de garanție pentru produsele destinate să aibă o folosință mai îndelungată. În mod normal, nu se vor prevedea

termene de garanție pentru pixuri, veselă etc. Aceste produse au, prin esența lor, o durată mai mult sau mai puțin scurtă. Durata de funcționare a veselei, de pildă, este în pericol, dacă spălatul ei este făcut de persoane mai puțin îndemînatice.

Dar, despre problemele de siguranță în funcționare vom vorbi mai tîrziu. Acum să revenim și să considerăm v o vîrstă oarecare inițială și fie $P(v)$ numărul de indivizi de vîrstă v ce compun populația a cărei mortalitate o investigăm. Valoarea medie a numărului de indivizi din mulțimea $\{P(v)\}$ care reușesc să împlinească x ani $(v < x)$ se notează cu $P(x)$ și poartă numele de *funcție de supraviețuire.*

Să vedem acum cum ajungem la $P(x + \mathrm{d}x)$ plecînd de la $P(x)$. Va trebui, prin urmare, să scădem din $P(x)$ numărul de indivizi care s-au stins în intervalul de timp $\mathrm{d}x$. Acest număr este însă $P(x)\,\lambda(x)\,\mathrm{d}x$. Așadar:

$$P(x + \mathrm{d}x) = P(x) - P(x)\,\lambda(x)\,\mathrm{d}x. \qquad (12)$$

care poate fi scrisă:

$$\lambda(x)\,\mathrm{d}x = -\frac{P(x + \mathrm{d}x) - P(x)}{P(x)}. \qquad (13)$$

Intervine acum o celebră teoremă din analiza matematică studiată în liceu: este teorema creșterilor finite sau teorema lui *Lagrange.* Funcția $P(x)$ satisface condițiile teoremei și deci putem scrie:

$$P(x + \mathrm{d}x) - P(x) = \mathrm{d}x P'(x). \qquad (14)$$

Așadar:

$$\lambda(x) = -\frac{P'(x)}{P(x)} \qquad (15)$$

sau încă:

$$\lambda(x) = -[\ln P(x)]' \qquad (16)$$

(trecerea se explică simplu dacă ținem cont de derivata unui logaritm).

Deci, rata mortalității este derivata logaritmică a funcției de supraviețuire luată cu semn schimbat.

Funcția de supraviețuire se deduce deci în funcție de rata mortalității prin integrare. Astfel:

$$P(x) = \mathrm{e}^{-\int \lambda(x)\,\mathrm{d}x + C} \qquad (17)$$

Să observăm, mai întîi, că funcția de supraviețuire trebuie să fie descrescătoare deoarece, în mod natural, odată cu trecerea timpului șansa de supraviețuire scade.

Prima problemă care se pune este aceea a determinării formei funcției $\lambda(x)$. Știm că ea trebuie să fie crescătoare; există însă o infinitate de funcții crescătoare. De exemplu:

$$f_n(x) = ax^n, \quad a > 0, \quad n \in N, \quad x > 0, \qquad (18)$$

unde a este un parametru pozitiv oarecare, iar n un număr natural. Or, dacă alegem succesiv:

$$f_1(x) = ax, \quad f_2(x) = ax^2, \ldots, f_n(x) = ax^n \qquad (19)$$

obținem o infinitate de funcții crescătoare din familia $f_n(x)$.

De-a lungul timpului s-a constatat însă că fenomenul de mortalitate — ca de altfel și cel al defectării sistemelor tehnice — are anumite particularități care ne dau posibilitatea să ne limităm la un număr rezonabil de clase de funcții adecvate pentru rata mortalității.

Totuși, problema nu este deloc de neglijat: cum se procedează? Ceea ce expunem aici constituie una din problemele de bază ale statisticii, de o deosebită importanță practică.

Avem în față un fenomen oarecare luat în studiu. Acest fenomen ne furnizează o serie de date de observație sau experimentale. Datele, la rîndul lor, au o anumită tendință de a se dispune într-un anume mod, conform fenomenului concret analizat. Ne găsim acum în fața dilemei: ce model statistic să alegem? Va trebui, bineînțeles, să facem o ipoteză. Această ipoteză prezintă o particularitate deosebită: ea nu este o ipoteză științifică (care caută să explice un fenomen, de exemplu, ipoteza structurii atomice a materiei etc.), ci o *ipoteză statistică,* adică este o presupunere asupra popula-

ției statistice, a colectivității luate în studiu. (Aceasta nu înseamnă că ipotezele statistice nu sînt făcute în concordanță cu principiile științifice. Aici cuvîntul „științific" nu este utilizat în mod absolut, ca antonim al cuvîntului neștiințific, adică mistic sau altcumva).

Ipoteza statistică este, în mod uzual, o *descriere concretă* a unui aspect sau a mai multor aspecte legate de populația respectivă.

Deși analiza concretă o efectuăm prin intermediul selecției, ipoteza statistică o vom face întotdeauna asupra populației, adică asupra întregului și nu asupra părții.

De ce această presupunere asupra unui aspect al populației este numită ipoteză statistică, este simplu de explicat, deoarece ea se referă la o situație care *poate* fi adevărată.

Acest *poate* pe care l-am subliniat, nu are deloc sens poetic; el descrie o situație reală care implică prin natura ei apariția unei probabilități.

Să căutăm însă să dăm un exemplu ca să înțelegem mai bine despre ce este vorba: fiindcă tot am pornit de la mortalitate, să ne menținem la ea. Să presupunem că alegem pentru $\lambda(x)$ (rata mortalității) următoarea funcție:

$$\lambda(x) = e^{\theta x+\omega}, \quad x > 0, \quad \theta, \omega > 0. \qquad (20)$$

Ea îndeplinește condițiile cerute mai înainte: este pozitivă, crescătoare în raport cu timpul (x), deci are toate calitățile unei intensități a mortalității.

Ceea ce am făcut deci în acest moment a fost avansarea unei ipoteze statistice asupra ratei mortalității la populația luată în studiu. Ceea ce avem noi la îndemînă sînt datele de observație, adică vîrstele la care s-au stins diverșii indivizi ai populației. Ceea ce trebuie să facem în continuare este să verificăm dacă această ipoteză statistică este adevărată, deoarece s-ar putea foarte bine să existe o altă funcție $\lambda(x)$ care să descrie rata mortalității, deci să existe *o alternativă* la ipoteza noastră inițială, numită în limbaj statistic *ipoteza nulă* (în toate limbile s-a încetățenit această denumire din

engleză, „null hypothesis" și notată H_0, alternativa fiind notată cu H_1).

Trebuie aici să insistăm mai mult, deoarece ne întîlnim în acest moment cu unul din punctele cheie ale gîndirii statistice.

În verificarea ipotezelor statistice sîntem întotdeauna puși în situația să comitem două genuri de erori:

1°. O eroare care constă în respingerea ipotezei H_0 atunci cînd ea este adevărată (numită și eroare de genul I) și care eroare dorim să o comitem cît mai rar. Din punct de vedere statistic, ea este o probabilitate și anume probabilitatea de a respinge H_0 atunci cînd în realitate H_0 este adevărată.

În mod uzual, această probabilitate se notează cu α și în practică se alege $\alpha = 0{,}01$ sau $\alpha = 0{,}05$;

2°. O eroare care constă în acceptarea ipotezei *Ho* cînd în realitate ea este falsă (numită și eroare de genul II).

Ea se notează în mod uzual cu β și în practică se alege $\beta = 0{,}10$ sau $\beta = 0{,}05$.

Care este însă procedura de verificare a unei ipoteze statistice?

Ea se face prin intermediul așa-numitelor *teste* (criterii) *statistice care sînt procedee statistico-matematice specifice științei statisticii.*

Nu trebuie să confundăm aici sensul obișnuit al cuvîntului „test" cu semnificația lui statistică. Testul statistic este, în final, un procedeu de calcul bazat pe anumite teoreme din statistica matematică prin care se ia pînă la urmă o decizie privind acceptarea sau respingerea ipotezei nule cu un anumit *nivel de încredere* dinainte fixat.

Ce este nivelul de încredere, ne putem ușor da seama, dacă vom considera probabilitatea $(1 - \alpha)$, care reprezintă probabilitatea de acceptare a ipotezei H_0 atunci cînd ea este adevărată.

Atunci, în mod analog, probabilitatea $(1 - \beta)$ o putem numi *puterea testului* respectiv, deoarece ea reprezintă probabilitatea de respingere a ipotezei H_0 cînd ea este în realitate falsă.

Să nu trecem însă peste problema pe care singuri ne-am pus-o și anume verificarea ipotezei statistice că rata mortalității are forma exponențială prezentată mai înainte. Testele construite pentru verificarea unor ipoteze statistice de acest tip poartă numele de „teste de concordanță" deoarece, de fapt, vrem să stabilim o concordanță între o repartiție empirică furnizată de datele de observație și o repartiție teoretică[1].

Ipotezele statistice referitoare la concordanță sînt ipoteze de un tip special. Am început însă cu ele deoarece par a fi, în mod logic, primul pas în analiza statistico-matematică a datelor de observație din orice domeniu, nu numai din cele legate de aspecte ale vieții sociale. Într-adevăr, stabilirea repartiției unui fenomen întîmplător (aleator) deschide ulterior calea pentru o analiză mai profundă, analiză care include predicții asupra dezvoltării viitoare a fenomenului respectiv.

„În general, pentru ca o distribuție[2] teoretică să aproximeze o distribuție empirică, trebuie ca frecvențele experimentale să nu difere mult de cele teoretice. Aceasta face ca și elementele caracteristice ale celor două distribuții să ia valori apropiate.

După cum am mai spus, graficul unei serii statistice poate sugera distribuția teoretică corespunzătoare. Un grafic sub formă aproximativă de clopot cu o pronunțată tendință spre simetrie sugerează o lege normală".

Frazele de mai sus sînt citate din manualul de clasa a XII-a liceu. Dacă citim cu atenție, observăm că în fond aici este vorba tocmai de problema noastră: stabilirea concordanței.

Desigur, manualul nu putea să trateze aceste probleme în detaliu, care depășeau materia prezentată în liceu totuși, la modul general, el ne avertizează de existența problemei și de importanța ei.

Problema concordanței a fost în mod general rezolvată încă la începutul acestui secol de către un clasic al statisticii britanice, Karl Pearson, care timp de mulți

[1] Vezi manualul de liceu, sau Craiu, V., *Verificarea ipotezelor statistice*, Editura didactică și pedagogică, București, 1972.

[2] Denumită și repartiție.

ani a condus revista „Biometrika", menționată mai înainte și care a fost una din primele reviste de statistică teoretică și aplicativă din lume.

Mai tîrziu, matematicienii sovietici A. N. Kolmogorov și N. V. Smirnov au elaborat un procedeu de verificare a concordanței care le poartă numele (testul Kolmogorov-Smirnov), și care este aplicabil pentru verificarea concordanței unei repartiții empirice cu orice repartiție teoretică considerată.

Testele Pearson și Kolmogorov-Smirnov sînt deci niște teste generale pentru stabilirea concordanței.

Deseori, în practică, ne interesează verificarea concordanței unei repartiții empirice cu o repartiție teoretică specificată cum ar fi, de exemplu, cea normală. Întrucît această repartiție se întîlnește foarte des, statisticienii s-au gîndit la construirea unor teste speciale care să verifice în mod specific legea normală.

Desigur, nimeni nu ne împiedică să aplicăm testele clasice, dar este de bănuit că din moment ce s-au construit teste speciale pentru o anume repartiție, atunci aceste teste speciale pot prezenta unele avantaje față de testele generale aplicate pentru aceeași repartiție.

Iată, de exemplu, să presupunem că datele de observație se dispun în mod pronunțat simetric în felul următor:

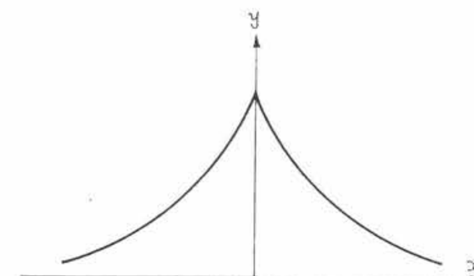


Fig. 7 — O dispunere posibilă a datelor de observație

Testele clasice s-ar putea să nu aibă suficientă putere de discriminare între două repartiții simetrice, deci s-ar putea să acceptăm ipoteza normalității (adică

repartiţia teoretică este normală) deşi în realitate ea nu este normală.

Aceasta nu înseamnă că testele clasice nu sînt bune, dar întotdeauna trebuie luaţi în considerare mai mulţi factori atunci cînd căutăm să rezolvăm o astfel de problemă de stabilire a concordanţei, cum ar fi:

1. Cîte date experimentale avem la dispoziţie?

Studii ulterioare asupra testelor clasice au evidenţiat faptul că acestea dau rezultate conforme cu realitatea dacă volumul selecţiei (n) deţinute este suficient de mare, în general $n \geqslant 100$;

2. Ce alternativă să alegem?

Într-adevăr, cineva ne poate spune: bine, facem ipoteza că repartiţia este normală deci H_0 este:

H_0: Funcţia de repartiţie este normală de medie „atît" şi dispersie „atît".

Dar atunci, care este alternativa? Că nu este normală, este foarte vag; vă puteţi imagina o mulţime destul de bogată de repartiţii ce nu sînt normale. Un exemplu e chiar cel din figura 7. Luăm atunci alternativa simetrică?

În mod firesc, această alternativă pare mai „normală" decît alternativa asimetrică, ca de exemplu:

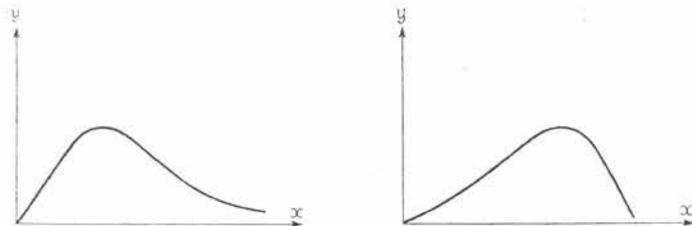


Fig. 8 — Repartiţii asimetrice cu „coadă" la dreapta sau la stînga

deoarece, dacă din însăşi natura fenomenului cît şi din reprezentarea grafică ne dăm seama că avem de-a face cu o repartiţie asimetrică, atunci bineînţeles că n-are nici un rost să presupunem normalitatea şi vom căuta să avansăm o altă repartiţie pentru a verifica concordanţa cu aceasta.

În fine, ar mai fi şi alţi factori; nu am vrut decît să evidenţiem că problemele de statistică nu sînt chiar atît de simple cum tind să fie considerate de unii dintre noi. Desprinderea de concret, de situaţia reală pe care o investigăm nu ne va conduce decît la aplicarea mecanică a unor procedee de calcul, a căror rezultate vor fi de cele mai multe ori în contradicţie cu realitatea.

Cu alte cuvinte, ştiinţa statisticii este foarte „pămînteană". Statisticianul englez M. G. Kendall, spunea, cîţiva ani în urmă, următoarele:

„Dar pentru matematicianul „purist" toate lucrurile sînt pure pe cînd statistica — să privim lucrurile în faţă — nu este o ştiinţă foarte „pură". Lucrăm prea aproape de problemele umanităţii"[1].

Cu toate aceste digresiuni, să revenim la problema noastră:

$$\lambda(x) = e^{\theta x + \omega}, \quad x > 0, \quad \theta, \omega > 0. \qquad (21)$$

Care este, în fond, problema?

1°. Să stabilim concordanţa între datele de mortalitate şi o funcţie de rată a mortalităţii de tipul de mai sus;

2°. Să estimăm parametrii θ şi ω care sînt, de fapt, necunoscutele.

După cele ce am spus, va trebui deci să aplicăm unul din „testele de concordanţă" despre care am discutat recent.

Chiar dacă nu ştim care este mecanismul lor concret de calcul, din însăşi citatul din manual ne dăm seama că avem nevoie de funcţia de repartiţie teoretică corespunzătoare lui $\lambda(x)$.

Să facem acest lucru: scriem din nou cine este în general expresia lui $\lambda(x)$ şi anume:

$$\lambda(x) = -\frac{P'(x)}{P(x)}. \qquad (22)$$

[1] Kendall, M. G., *Statistics and Management*, În: „Proc. First Canad. Conf. in Applied Statistics „Statistics '71, Canada", Mc Gill Univ., 1971, p. 13.

Dar funcția de supraviețuire $P(x)$ înseamnă, în fond, probabilitatea ca un individ luat la întîmplare să trăiască cel puțin x ani, adică $P(x) = \text{Prob}\{X \geqslant x\}$.

Atunci, $1 - P(x) = \text{Prob}\{X < x\}$, iar această probabilitate este tocmai funcția de repartiție:

$$F(x) = \text{Prob}\{X < x\}. \tag{23}$$

Așadar, derivata cu semn schimbat a funcției de supraviețuire este tocmai $F(x)$. Într-adevăr:

$$-P'(x) = [1 - P(x)]' = [F(x)]'. \tag{24}$$

Dar această ultimă derivată (a funcției de repartiție) ne dă tocmai densitatea de probabilitate, notată $f(x)$,

$$F'(x) = f(x). \tag{25}$$

Reamintim aici proprietățile unei densități de probabilitate:

1°. $f(x) \geqslant 0$ pentru $x \in D$, D fiind domeniul de variație al variabilei aleatoare considerate;

2°. $\int_D f(x)\,\mathrm{d}x = 1$.

În general, se consideră $D \equiv (0, \infty)$ adică $x > 0$, fiindcă durata de viață (atît a noastră cît și a obiectelor ce ne înconjoară) este pozitivă.

Faptul că se consideră semidreapta drept domeniu, adică la celălalt capăt al intervalului se consideră infinit, poate suscita comentarii, atît în teoria mortalității, cît și în teoria siguranței sistemelor tehnice.

Într-adevăr, viața noastră este finită și, de asemenea, și durata de funcționare a majorității sistemelor tehnice de care ne folosim.

Astfel, în teoria mortalității se poate lua, să zicem, $D \equiv (0, 300)$, 300 reprezentînd ani, pentru a acoperi cu probabilitatea 1 orice excepție biologică. Pînă în prezent, nu pare deloc absurd să presupunem că fiecare dintre noi va înceta din viață la o vîrstă ceva mai mică de 300 de ani ...

Pentru obiecte care au o durată de viață mai mare, putem lua marginea superioară a intervalului oricît de mare vrem, sau cel puțin cît ni se pare rezonabil.

În acest caz, presupunerea domeniului $(0, \infty)$ nu apare nenaturală.

Presupunerea cea mai conformă cu realitatea ar fi aceea să alegem intervalul de forma $(0, b)$, unde b este un parametru necunoscut, care va trebui determinat pentru fiecare repartiție presupusă în parte.

Problema nu este însă chiar atît de simplă. Faptul că definim variabila aleatoare pe intervalul $(0, b)$ înseamnă că densitatea ei trebuie să satisfacă relația:

$$\int_0^b f(x)\,\mathrm{d}x = 1. \tag{26}$$

Or, aceasta implică apariția parametrului b în forma densității. Iată un exemplu:

$$f(x;\,\delta, b) = \delta b^{-\delta} x^{\delta-1} \quad 0 < x < b, \quad \delta > 0. \tag{27}$$

Nu este greu de văzut că funcția de repartiție corespunzătoare (adică integrala lui $f(x)$) este:

$$F(x;\ \delta, b) = \left(\frac{x}{b}\right)^{\delta}, \quad 0 < x < b, \quad \delta > 0. \tag{28}$$

Această repartiție poartă numele de „repartiția putere" (engl. „power distribution"; pol. „rozklad potegowy") și a fost studiată pe larg de doi ingineri polonezi[1] de la Institutul de automatică al Academiei R. P. Polone, ca model pentru fenomenul de defectare a unor componente electrice ale sistemelor tehnice folosite în automatizarea proceselor tehnologice.

Iată deci una din problemele deschise în statistica matematică, în care mai sînt multe lucruri de făcut.

În fine, am ajuns la concluzia că $\lambda(x)$ se mai poate scrie:

$$\lambda(x) = \frac{f(x)}{1 - F(x)}. \tag{29}$$

Din această formulă, ne dăm seama, poate mai bine, de semnificația practică a lui $\lambda(x)$: raportul dintre numărul indivizilor ce s-au stins pînă la momentul x și al celor ce au supraviețuit pînă la același moment.

[1] Firkowicz, Sz. și Ciechanowicz, K.

Astfel, se constată că intensitatea sau rata mortalității se poate „estima" din datele de observație, deoarece, atît momentele (x), cît și numărul indivizilor dintr-o categorie sau alta, este cunoscut.

În acest mod, se obține o rată a mortalității empirică.

Relația de mai sus se scrie, ținînd cont că $f(x) = F'(x)$, astfel:

$$F'(x) + \lambda(x) F(x) = \lambda(x), \qquad (30)$$

care este o ecuație diferențială simplă ce ne furnizează pe $F(x)$ — funcția de repartiție în cazul în care cunoaștem pe $\lambda(x)$.

Un mic calcul ne dă densitatea de probabilitate sub forma:

$$f(x) = \lambda(x) \cdot e^{-\int_0^x \lambda(u)\,du}. \qquad (31)$$

Pentru a vedea ce simplu este calculul, să deducem relația de mai sus, astfel:

Am găsit că:

$$P(x) = e^{-\int_0^x \lambda(u)\,du}, \qquad (32)$$

dar din definiție, $\lambda(x)$ se mai scrie

$$\lambda(x) = \frac{f(x)}{P(x)}, \quad \text{deci } f(x) = \lambda(x)\, P(x) \qquad (33)$$

deci, banal, după o expresie favorită a matematicienilor.

Dacă modul de calcul e banal, nu și banală este informația obținută cu ajutorul acestei relații.

Cunoscînd pe $\lambda(x)$ deducem pe $f(x)$ și în continuare pe $F(x)$.

Acum putem, în fine, să începem procedura de stabilire a concordanței.

Avem funcția teoretică de repartiție corespunzătoare unui $\lambda(x)$ ales, din datele de observație cunoaștem repartiția empirică deci totul se reduce la aplicarea testului Kolmogorov-Smirnov.

Dar, deși calculul propriu-zis legat de acest test este foarte simplu — el reducîndu-se, în esență, la operații cu numere — totuși teorema care stă la baza acestui test necesită un aparat matematic destul de complicat.

Bineînțeles că practicianul din domeniul științelor sociale, ca de altfel din orice alt domeniu, nu are nevoie să cunoască demonstrația teoremei și toate celelalte subtilități matematice legate de aceasta. El trebuie să fie în posesia instrumentului practic de lucru, să știe cînd, cum și unde să-l aplice, în funcție de condițiile reale ale problemei respective, să știe cum să interpreteze rezultatele obținute, iar pe baza acestora să ia o decizie. În acest din urmă pas — luarea deciziei —, statistica înmînează ștafeta domeniului specific. Intervine acum gîndirea specifică domeniului social, economic sau tehnic.

Desigur că nu vom prezenta aici acest test: lăsăm cititorului curiozitatea de a răsfoi o carte de statistică matematică pentru a vedea cam în ce mod se pun problemele.

Totuși, fiindcă atît am purtat cu noi pe $\lambda(x)$ nu-l vom părăsi de tot.

Ca informație istorică, acest $\lambda(x)$ sub forma exponențială a fost propus în 1825 de W. Gompertz într-un articol — azi celebru — privind teoria mortalității. Desigur, el nu avea la îndemînă aparatul statistico-matematic modern de analiză a datelor experimentale, dar din studierea aprofundată a fenomenului de mortalitate el a presupus în mod corect că una din reprezentările posibile ale ratei mortalității poate fi și cea exponențială $(e^{\theta x+\omega})$.

Repartiția generată de $\lambda(x) = e^{\theta x+\omega}$ poartă și numele de „repartiția Gompertz".

Dacă pentru o populație a cărei mortalitate o studiem, rata mortalității este de tip Gompertz, atunci luînd logaritmul natural obținem:

$$\ln \lambda(x) = \theta x + \omega. \qquad (34)$$

Să notăm, pentru simplificare $y = \ln \lambda(x)$ și astfel obținem relația

$$y = \theta x + \omega. \quad (35)$$

Presupunem că $x_1, x_2, ..., x_n$ sînt vîrstele (anii) la care se sting diferiții indivizi ai populației. Mulțimea de numere $\{x_1, x_2, ..., x_n\}$ formează deci selecția noastră sau „seria statistică", așa cum se studiază în liceu.

Pe de altă parte, mulțimea de numere $\{\lambda(x_1), ..., ..., \lambda(x_n)\}$ reprezintă ratele de mortalitate estimate pentru fiecare

$$x_i, \quad i = 1, 2, ..., n.$$

Dacă vom reprezenta într-un sistem de axe rectangular — pe axa Oy, valorile y_i, $i = 1, 2, ..., n$, iar pe axa Ox datele de observație x_i, $i = 1, 2, ..., n$ — atunci, dacă rata mortalității (vorbim de rata teoretică acum, de modelul statistic în sine) este de tip Gompertz, punctele de coordonare (x_i, y_i) $i = 1, 2, ..., n$ trebuie să se dispună aproximativ după o dreaptă:

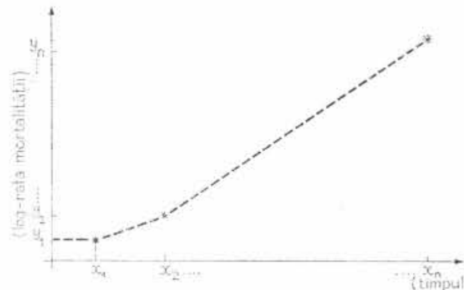


Fig. 9 — Ajustarea unei curbe de mortalitate de tip Gompertz prin transformare logaritmică

Ceea ce ne mai rămîne de făcut este să determinăm dreapta corespunzătoare, adică să estimăm parametrii θ și ω.

Metoda de determinare a acestei drepte se numește *ajustarea liniară*, iar procedeul efectiv este dat de așa-numita *metodă a celor mai mici pătrate*. În cazul liniar,

această metodă este simplă, iar logica ei e, de asemenea, ușor de înțeles.

Să introducem în ecuația dreptei $y = \theta x + \omega$, pe rînd, valorile x_i și y_i și să le privim în paralel:

y_i	$\theta x_i + \omega$
Este valoarea estimată din datele de observație, $y_i = \ln \lambda(x_i)$	Este valoarea calculată a lui y_i prin ecuația dreptei $y = \theta x + \omega$ în punctul x_i

Este atunci clar că diferența:

$$y_i - (\theta x_i + \omega),$$

notată în mod uzual „e_i", reprezintă tocmai eroarea dintre valoarea reală (y_i) obținută din practică și valoarea calculată $(\theta x_i + \omega)$ din presupunerea teoretică.

Ajustarea este cu atît mai bună cu cît eroarea totală va fi mai mică.

Iată deci că am ajuns la o altă problemă de analiză matematică și anume la găsirea unui minim.

S-ar putea cere, de exemplu, ca suma modulelor erorilor să fie minimă, adică:

$$\sum_{i=1}^{n} |e_i| = \text{minim}. \quad (36)$$

Dar, se știe din analiză că modulul unei funcții nu are derivată în orice punct și deci acest mod de rezolvare nu este practic.

Gauss, care este unul dintre matematicienii [1] ce s-au ocupat intens de această metodă, a celor mai mici

[1] Primele elemente i se datorează lui Legendre, care, în 1806, le-a aplicat în probleme de astronomie, la calculul orbitelor unor comete.

pătrate, a arătat că dacă se consideră suma pătratelor erorilor:

$$\zeta = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - (\theta x_i + \omega)]^2 \text{ (în cazul nostru)}, \quad (37)$$

atunci minimul lui ζ există și el se obține ca soluție a sistemului de ecuații:

$$\frac{\partial \zeta}{\partial \theta} = 0, \quad \frac{\partial \zeta}{\partial \omega} = 0, \quad (38)$$

numit și „sistemul de ecuații normale".

Metoda lui Gauss este generală, adică dacă avem, în general, o legătură funcțională φ a cărei formă o cunoaștem:

$$y = \varphi(x;\ \theta_1, \theta_2, ..., \theta_k), \quad (39)$$

unde necunoscutele sînt $\theta_1, \theta_2, ..., \theta_k$, iar x și y se cunosc prin observare, măsurare sau orice alt procedeu, atunci formînd suma pătratelor erorilor:

$$\zeta = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} [y_i - \varphi(x_i;\ \theta_1, \theta_2, ..., \theta_k)]^2 \quad (40)$$

atunci sistemul de ecuații normale:

$$\frac{\partial \zeta}{\partial \theta_j} = 0, \quad j = 1, 2, ..., k \quad (41)$$

ne furnizează minimul dorit.

Necunoscutele fiind θ_j, ele se deduc deci din rezolvarea sistemului de mai sus, obținîndu-se ca soluții în funcție de x_i și y_i, care sînt datele de observație.

Să aplicăm aceasta în cazul liniar; derivăm întîi în funcție de θ. Avem:

$$\frac{\partial \zeta}{\partial \theta} = -2 \sum_{i=1}^{n} x_i [y_i - (\theta x_i + \omega)] = 0 \quad (42)$$

sau:

$$\left(\sum_{i=1}^{n} x_i^2\right) \theta + \left(\sum_{i=1}^{n} x_i\right) \omega = \sum_{i=1}^{n} x_i y_i. \quad (43)$$

Derivăm acum în funcție de ω. Avem:

$$\frac{\partial \zeta}{\partial \omega} = -2 \sum_{i=1}^{n} [y_i - (\theta x_i + \omega)] = 0 \quad (44)$$

sau:

$$\left(\sum_{i=1}^{n} x_i\right) \theta + n\omega = \sum_{i=1}^{n} y_i. \quad (45)$$

Așadar, în final, avem de rezolvat sistemul liniar în necunoscutele θ și ω:

$$\begin{cases} n\omega + (\Sigma x_i)\,\theta = \Sigma y_i \\ (\Sigma x_i)\,\omega + (\Sigma x_i^2)\,\theta = \Sigma x_i y_i \end{cases} \quad (46)$$

care rezolvare este, evident, foarte simplă. Nu-l mai rezolvăm aici. Spunem doar că soluțiile se notează cu $\hat{\theta}$, $\hat{\omega}$ și ele se numesc „estimațiile prin metoda celor mai mici pătrate" ale parametrilor θ, ω.

Metoda celor mai mici pătrate o vom găsi aplicată în cele mai diverse domenii. După cum ați putut observa, în cazul legăturilor simple, cum ar fi cea liniară, ea este și ușor de manipulat, necesitînd doar atenție la calcul și, bineînțeles, asimilarea cunoștințelor de matematică ce se predau în licee.

Am încercat pînă aici să vă înfățișăm cîteva aspecte legate de gîndirea statistică, de metodele statistice în domeniul vieții sociale. Poate că domeniul specific ales și anume mortalitatea nu este prea îmbietor pentru cineva care face acum cunoștință cu statistica, dar așa s-au petrecut lucrurile din punct de vedere istoric.

Pe de altă parte, teoria mortalității și-a găsit, ulterior, în epoca modernă, o aplicație de importanță vitală pentru societate: teoria siguranței în funcționare a sistemelor tehnice. Despre importanța acesteia nu este nevoie să argumentăm. Să ne gîndim numai la problemele ridicate de siguranța în funcționare a sistemelor folosite la zborurile cosmice.

Nu știu dacă cele discutate pînă acum v-au trezit curiozitatea pentru teoria mortalității și mai general către problemele de demografie. În orice caz, n-ar fi deloc inutil să deschideți o dată, de curiozitate, *Anuarul demografic al R. S. România* — să zicem cel din 1974 —

publicație editată de Direcția Centrală de Statistică și de Comisia Națională de Demografie.

Veți găsi aici lucruri foarte interesante privind diferite aspecte legate de populația țării noastre ca, de exemplu:

— structura demografică și social-economică a populației;

— mișcarea naturală a populației (nașteri, decese, căsătorii, divorțuri);

— mișcarea migratoare internă a populației.

Putem astfel veni în contact cu datele numerice ce reflectă diferite fenomene pe care le cunoaștem la modul general: scăderea mortalității infantile, creșterea mediei de viață, scăderea numărului divorțurilor etc.

Mai putem afla: ce lună deține recordul de căsătorii încheiate de-a lungul mai multor ani, cauzele principale de deces etc.

Multe din aceste aspecte sînt ilustrate cu grafice care constituie, printre altele, o bună aplicație a celor ce știm pînă acum despre prezentarea datelor statistice. Cu unele din astfel de probleme ne vom întîlni însă în subparagraful următor.

B. Recensămînturile

Aplicațiile matematice și statistico-matematice care se introduc tot mai mult în diferite ramuri ale economiei naționale, învederează un pas important în conducerea științifică ...

Gh. Mihoc[1]

Problema recensămîntului este foarte veche. De îndată ce societatea umană s-a organizat în formații statale, mai mult sau mai puțin evoluate, problema

[1] Mihoc, Gh., *Rolul statisticii matematice în realizarea obiectivelor economice.* În: „Metode noi și probleme de perspectivă ale cercetării științifice", Editura Academiei R. S. România, 1970, p. 269.

„evidenței" în diferite ramuri de activitate s-a pus în mod natural. După cum e de imaginat, și haosul în această evidență era foarte mare. O parte din cauzele acestui haos le-am mai discutat înainte.

Pe de altă parte, nici nu putem pretinde ca recensămînturile efectuate acum cîteva secole să fie la fel de complete și cuprinzătoare, sub toate aspectele, așa cum sînt cele din epoca noastră, care au la dispoziție întregul arsenal atît material, cît și teoretic pentru ducerea la îndeplinire a unei astfel de activități.

De la bun început, trebuie să ne fie clar că prin recensămînt nu înțelegem numai înregistrarea ca atare a indivizilor dintr-o anumită zonă geografică. Într-un recensămînt — pe lîngă numărul indivizilor — interesează o sumedenie de aspecte legate de structura mulțimii indivizilor, de diferite submulțimi ale ei, clasificări în cadrul submulțimilor, relații, comparații, sau interdependențe între aceste submulțimi etc. În acest etc. sînt incluse aspecte economico-sociale, ecologice (de mediu) și, de asemenea, eventuale previziuni pe baza datelor sintetizate într-un anumit moment de timp.

Deși limbajul întrebuințat de noi aici este oarecum vag, ne dăm seama că un recensămînt reprezintă o acțiune complexă, care necesită o mare concentrare de forțe pentru efectuarea ei. Din această cauză, recensămînturile nu se fac prea des.[1]

Pentru a lămuri mai bine cum stau lucrurile cu problema recensămînturilor, să urmărim — tot istoric — evoluția și caracteristicile recensămînturilor efectuate în țara noastră.

Primul recensămînt al populației de pe teritoriul țării noastre a fost inițiat în anul 1838 de către așa-numitul „Departament al Treburilor din Lăuntru" al Țării Românești. Alături de date privind populația ca atare, s-au luat în considerare și alte elemente legate de industrie, agricultură, locuințe. Din această cauză, re-

[1] Pe baza recensămînturilor precum și a datelor colectate anual asupra populației se alcătuiesc *anuarele demografice naționale.*

censămîntul din 1838 se poate considera ca primul nostru recensămînt, în sensul modern.

Următoarele recensăminturi au avut loc după cum urmează:

- — al doilea: decembrie 1859 — martie 1860,
- — al treilea: 1899,
- — al patrulea: 1912,
- — al cincilea: 1930,
- — al şaselea: 1941 (aprilie),
- — al şaptelea: 1948 (ianuarie),
- — al optulea: 1956 (februarie),
- — al nouălea: 1966 (martie),
- — al zecelea: 1977 (ianuarie).

Observăm o periodicitate constantă (aproximativ un deceniu) de efectuare a recensămînturilor în anii de după Eliberare.

Este foarte interesant să urmărim care au fost caracteristicile, obiectul recensămînturilor respective. În tabelul de mai jos, adaptat dintr-o lucrare foarte interesantă privind populaţia României, se indică aceste caracteristici:

Ce a interesat în recensămînturile efectuate în România	Anii în care au fost efectuate recensămînturile								
	1838	1859—60	1899	1912	1930	1941	1948	1956	1966
Populaţia	da	da	da	da	da	da	da	da	da
Locuinţele	da	—	—	da	da	da	da	—	da
Clădirile de locuit	da	—	—	da	da	da	da	—	da
Industria	da	—	—	—	da	da	—	—	—
Agricultura	da	—	—	—	—	da	da	—	—

Tabelul 7. Obiectivele recesămînturilor efectuate pînă în 1974 în România[1].

[1] *Populaţia României* (coordonatori Retegan, G., Păcurariu, I.), Edit. Meridiane, Bucureşti, 1974, p. 90.

Iată un bun prilej pentru a desfăşura unele comentarii asupra acestui tabel statistic „calitativ". L-am numit astfel, deoarece el nu conţine numere, ci indicaţii calitative.

Mai întîi, să înşirăm anii ce separă succesiv cele nouă recensăminturi româneşti:

22 — 39 — 13 — 18 — 11 — 7 — 8 — 10.

Dacă vom aduna intervalele primelor cinci recensăminturi, obţinem 103 ani. Aşadar, în decurs de un secol s-au efectuat la noi numai 6 recensăminturi la diferenţe de timp destul de neregulate.

Primul recensămînt — datorită lipsei de informaţie cvasitotală asupra situaţiei demo-economice a ţării — a avut drept obiective toate caracteristicile înscrise în prima coloană a tabelului.

Privind tabelul, se constată că, din nou, totalitatea obiectivelor este atinsă abia la 103 ani diferenţă, deci la recensămîntul din 1941.

În ceea ce priveşte recensămînturile din anii de după Eliberare, vă puneţi poate întrebarea de ce în 1948 nu s-a luat în considerare industria, iar în ceilalţi ani de ce s-au considerat doar anumite obiective, relativ puţine în comparaţie, să zicem, cu primul recensămînt — cel din secolul trecut.

Pentru recensămîntul din 1948 explicaţia este relativ simplă. Efectele negative ale războiului încă erau simţite — multe obiective economice trebuiau reconstruite, industrializarea ţării era încă în fază incipientă mai ales că recensămîntul a avut loc înainte de naţionalizarea mijloacelor de producţie (11 iunie 1948).

În ceea ce priveşte obiectivele — relativ limitate — ale unor recensămînturi ulterioare, justificarea nu mai

este chiar atît de evidentă. Aici, trebuie să ținem seama de faptul că, între timp, în țara noastră s-a introdus *sistemul unic de evidență națională* — o modalitate foarte eficientă de sintetizare a datelor statistice necesare acumulării unei informații cît mai complete privind evoluția economico-socială a țării. De asemenea, a luat ființă organismul specializat de stat în colectarea și prelucrarea informației statistice — Direcția centrală de statistică — cu filialele ei județene.

În același timp, o activitate organizată în direcția muncii statistice s-a elaborat și în cadrul diferitelor ramuri economice de profil în așa fel încît astăzi obiectivele unui recensămînt pot fi restrînse numai la caracteristicile necercetate prin mecanismele de evidență curentă.

Bineînțeles, obiectivele unui recensămînt pot fi alese — și se aleg — în funcție de necesitățile perioadei istorice respective, de gradul de standardizare atins pe plan mondial, de noile achiziții în domeniul teoriei și practicii statisticii demografice.

Un recensămînt nu este un lucru ușor de făcut: din tabel putem să tragem concluzia, la o primă vedere, că există doar 5 caracteristici generale; să ne gîndim însă la multiplele aspecte legate de industrie.

Chiar în cazul — să zicem mai simplu — al culegerii datelor privind populația efectivă, apar probleme pe care la început poate că le ignorăm. De exemplu: dacă dorim ca un recensămînt privind populația să fie util, trebuie să „intervievăm" fiecare persoană. În limbaj de specialitate se spune că unitatea de înregistrare este persoana. Acum vine problema cea mai dificilă a unui recensămînt și anume modul de înregistrare. Prin aceasta se înțelege felul în care se „intervievează" fiecare persoană: „prin interogare" sau prin „autoînregistrare" Mai explicit: prin autoînregistrare înseamnă, de pildă, să trimitem fiecărui cetățean al țării o scrisoare conținînd un formular cu caracteristicile ce ne interesează: demografice (sex, vîrstă, stare civilă etc.), sociale (naționalitate, nivel de instruire, grupa socială etc.), economice (profesia, locul de muncă etc.), urmînd ca fiecare cetățean să ne răspundă pînă la o dată fixată.

Dar, să fim realiști: de cîte ori noi înșine am răspuns unui prieten sau coleg la o scrisoare, imediat după ce am citit scrisoarea? Chiar dacă sîntem pătrunși de simțul datoriei, factori subiectivi și obiectivi ne fac uneori să nu răspundem în timp util unei astfel de acțiuni.

Într-adevăr, practica recensămînturilor pe plan mondial a evidențiat faptul că metoda autoînregistrării nu dă rezultatele scontate, tocmai pentru că dilată foarte mult timpul necesar sintetizării datelor respective. La noi, această practică, prin autoînregistrare, nu a fost aplicată decît în 1899, ulterior renunțîndu-se la ea în efectuarea recensămînturilor.

Nu trebuie să tragem concluzia că un astfel de procedeu este total nepractic; întotdeauna trebuie să se țină seama de scopurile pe care le urmărim. Un „chestionar" de genul celui întrebuințat la un recensămînt este, prin natura lui, mai complicat, în sensul că are multe puncte la care trebuie să se răspundă foarte exact.

În fine, nu vom intra prea mult în detaliile legate de chestionare în investigația statistică din domeniul sociologic. [1]

Trebuie să subliniem faptul că recensămînturile reprezintă una din modalitățile pe care își bazează acțiunile administrația de stat. Dacă, de exemplu, un recensămînt arată o carență într-un anumit domeniu al vieții economico-sociale, atunci corelînd datele obținute cu necesitățile și posibilitățile în acel domeniu, se pot lua măsuri practice pentru remedierea situației.

[1] Vezi: Chelcea, S, *Chestionarul în investigația sociologică*, Edit. științifică și enciclopedică, București, 1975.

Considerînd un domeniu foarte important din viața oricărei țări și anume problema populației, să privim, de exemplu, diagramele circulare care descriu structura populației pe grupe mari de vîrstă la recensămînturile din anii 1930 și 1956:

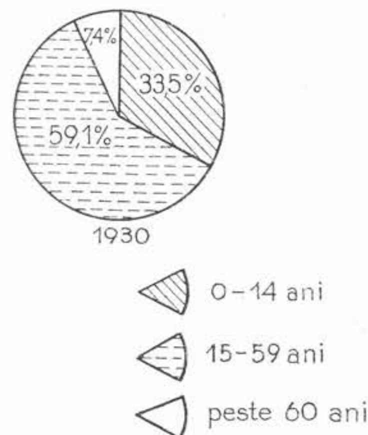


Fig. 10 — Structura populației pe grupe mari de vîrstă la recensămînturile din anii 1930 și 1956, în România

O simplă privire pe aceste diagrame ne face să tragem următoarele concluzii: deși ponderea populației apte de o activitate productivă (grupă 15—59 ani) a crescut în 1956 față de 1930 totuși se constată, în același timp, o mărire a grupei pasive (peste 60 ani) și o scădere a „fondului" demografic al națiunii — copiii.

O informație suplimentară, pe care o putem afla din analele demografice, este aceea că între anii 1930—1956 populația țării noastre a crescut, în medie, cu 1 pînă la 1,3% pe an.

În special, după anul 1948, cînd s-a instalat un nou regim de asistență medicală, mortalitatea generală a populației a scăzut considerabil. Acest fapt a avut drept consecință înlăturarea influenței negative a mortalității asupra creșterii populației țării.

În mod logic, rezultă că această creștere a populației va depinde acum aproape exclusiv de natalitate. La un moment dat, s-a constatat o scădere a natalității în perioada 1956—1966. Acest lucru se poate observa ușor, luînd în considerare populația între 0—14 ani (în special, contingentele 0—2 și 3—5 ani) în perioada amintită (vezi tabelul 8).

Grupa de vîrstă (ani)	Mii persoane		În % față de populația totală	
	1956	1966	1956	1966
0— 2	1126,1	812,7	6,4	4,2
3— 5	1106,4	894,7	6,3	4,7
6—14	2581,1	3261,1	14,8	17,1
Total	4813,6	4968,5	27,5	26,0

Tabelul 8: Populația din grupa 0—14 ani la recensămînturile din 1956 și 1966[1]

Analiza situației puse în evidență de aceste recensăminturi a arătat o scădere exagerată a natalității, care dacă ar fi continuat pe o perioadă de timp mai lungă ar fi pus în pericol resursele de muncă ale țării.

Ceea ce s-a petrecut între anii 1930—1966 este cunoscut sub numele de „fenomenul de îmbătrînire demografică". (În „Revista de Statistică", 1972, nr. 4, Vl. Trebici semnează un *Dicționar demografic multilingv,* unde se găsesc explicate principalele noțiuni de demografie. Același autor semnează în revista „Viitorul Social", nr. 1, 1972, un articol consacrat acestei probleme. Titlul este ***Îmbătrînirea demografică a populației***

[1] *Populația României,* op. cit., p. 31.

României. Vezi și *Mică enciclopedie de demografie*, Edit. științifică și enciclopedică, București, 1975).

Fenomenul de îmbătrînire demografică s-a petrecut și se petrece în multe țări de pe glob.

În Europa, acest proces a avut loc foarte rapid. Din tabelul de mai jos, putem vedea care este ponderea populației în vîrstă de 60 ani și peste, într-o serie de țări europene:

Țara	Populația în vîrstă de 60 ani (în procente)	Țara	Populația în vîrstă de 60 ani (în procente)
Turcia	7,2	Cehoslovacia	15,9
Albania	8,9	Ungaria	15,9
U.R.S.S.	9,8	Elveția	16,5
Iugoslavia	11,0	Danemarca	16,9
Polonia	12,1	Norvegia	17,1
Finlanda	12,3	Franța	18,0
Portugalia	12,4	Anglia	18,2
România	13,0	R. F. Germania	18,4
Bulgaria	13,4	Suedia	18,6
Grecia	13,8	Belgia	18,8
Olanda	14,0	R. D. Germană	22,0
Italia	14,5		

Tabelul 9: Ponderea populației în vîrstă de 60 ani și peste, într-o serie de țări europene, în anul 1970. (Sursa: Vl. Trebici, *Populația mondială*, Ed. științifică, București, 1974, p. 96.)

Din acest tabel se desprinde concluzia că, practic, doar Turcia prezintă un procent din categoria respectivă care să-i permită să considere că populația ei este tînără. O altă concluzie este aceea că îmbătrînirea demografică afectează cu deosebire țările puternic industrializate.

Nu întotdeauna fenomenul invers — adică cel de întinerire demografică a populației — apare și se menține pe cale naturală. Din această cauză, este necesară intervenția statului, care, printr-o politică adecvată a populației, să preîntîmpine efectele negative ale fenomenului.

Măsurile luate de statul nostru pentru întărirea familiei și creșterea natalității sînt în concordanță cu legile obiective de evoluție ale societății, cu grija pentru viitorul țării.

O altă problemă interesantă, pusă în evidență de datele statistice sintetizate în *Anuarul demografic al R. S. România* din 1974, este aceea a creșterii longevității în țara noastră. Acest fapt este, în primul rînd, o consecință a creșterii nivelului de trai general al populației, care, încă din primii ani după Eliberare, a avut ca efect imediat scăderea substanțială a mortalității infantile.

Astăzi, țara noastră a intrat în grupa țărilor pentru care durata medie de viață este de aproximativ 70 de ani, ceea ce reprezintă o realizare însemnată din punct de vedere social.

După cum se știe, durata medie de viață ridicată este caracteristică țărilor industrializate. Să privim acum tabelul 10, unde s-au cules datele referitoare la ponderea populației de 75 de ani și peste, în R. S. România, pentru a observa creșterea continuă a acestor contingente de vîrstă.

Durata medie de viață ridicată nu este dată însă — nici la noi, nici în alte părți — de numărul foarte mare de longevivi (persoane în vîrstă de 90 ani și peste), ci de numărul mare de persoane care pășesc „în a treia tinerețe" (50—70 ani). După cum este iarăși cunoscut, numărul mare de longevivi care au depășit

un secol de existență se întîlnește de multe ori în locuri unde nivelul de trai nu este prea înalt (de exemplu, în regiunea Anzilor) sau în locuri unde condițiile geografice și alimentația locuitorilor sînt oarecum ieșite din comun. (de exemplu, în regiunile muntoase ale Azerbaidjanului). Această problemă implică factori foarte complecși și putem spune că vîrstele longevivilor constituie pentru statistică un material concret pentru „teoria valorilor extreme".

Anul	Bărbați	Femei
1956	112462	175940
1957	123094	191696
1958	131663	203290
1959	137264	209426
1960	145113	221855
1961	153686	236113
1962	155405	240382
1963	158905	249015
1964	165979	261496
1965	171969	270261
1966	172429	284027
1967	181050	285863
1968	188177	295119
1969	196945	304329
1970	210317	313899

Tabelul 10: Numărul de persoane în vîrstă de 75 de ani și peste în perioada 1956–1970[1]

[1] Datele au fost luate din *Anuarul demografic al R. S. România*, 1974.

C. Sondarea opiniei publice

„Orice analiză statistică pornește de la un model probabilistic, care este construit cu scopul de a reprezenta aspectele cele mai importante ale structurii datelor."

W. G. Cochran[1]

Astăzi sîntem destul de obișnuiți cu așa-numitele „anchete sociale", care privesc unul sau altul din aspectele noastre de zi cu zi. Marile magazine ne înmînează formulare prin care ne cer părerea asupra unui anumit produs, juriul unui anumit concurs cere și părerea publicului despre participanți etc.

Toate acestea pot fi încadrate într-o acțiune mai generală, cunoscută azi sub numele de „sondaj al opiniei publice". Evident, acest „sondaj" (însăși termenul este statistic) nu se poate face pe întreaga populație, ci doar pe o parte a ei, adică pe o selecție.

În mod inevitabil, statistica cu metodele ei specifice intervine din nou. Istoria sondajelor de opinie publică începe în S.U.A. — nu cu o problemă economică, ci cu una politico-socială și anume cu sondarea opiniei publice privind alegerile prezidențiale.

Începînd din anul 1920 și apoi trecînd prin anii 1924, 1928 și 1932, s-au efectuat încercări privind „prezicerea" rezultatelor alegerilor în această țară. În 1932, revista „The Literary Digest" a prezis, cu o eroare mai mică de 1%, rezultatul alegerilor din 1932, ceea ce a însemnat un succes destul de mare și respectiv atragerea atenției publicului asupra acestor sondaje.

Anul 1936 a însemnat însă anul de grație pentru susamintita revistă (preludiul desființării ei!). În acea perioadă, lupta electorală s-a purtat între Franklin

[1] Cochran, W. G., *The potential contribution of electronic machines to the field of statistics*. În: „The Annals of Computation Laboratory of Harvard Univ.," vol. XXXI, 1962, p. 236.

D. Roosevelt și Alfred D. Landon. „The Literary Digest" a anunțat următorul sondaj de opinie:

CARACTERISTICA	Roosevelt	Landon
Numărul de state favorabile	16	32
Numărul de circumscripții electorale favorabile	161	370

Tabelul 11: Predicția pentru alegerile prezidențiale în 1936 în S.U.A. conform cu „The Literary Digest"[1]

După cum se știe, din evoluția ulterioară a evenimentelor istorice, victoria a repurtat-o F. D. Roosevelt iar diferențele dintre cifrele prezise și cele reale au fost considerabile:

Rezultatul	Prezis	Real
În procente	42,9	62,5
Număr circumscripții	161	523
Număr state	16	46

Tabelul 12: Compararea rezultatelor reale și prezise în cazul președintelui F.D. Roosevelt.

În mod normal, oricine se întreabă care este cauza unei astfel de predicții eronate. Pentru a putea afla un

[1] Anderson, T. W., Sclove, S. C., *Introductory Statistical Analysis*, Houghton Mifflin Co., Boston, 1974, p. 5.

răspuns, este foarte util să examinăm procedeul folosit de numita revistă pentru a trage unele concluzii asupra modului în care trebuie conduse sondajele de opinie publică, astfel încît eroarea să nu fie catastrofală.

Revista și-a propus să facă sondajul de opinie a unui număr de 10 milioane de alegători. În acest scop, s-a folosit metoda corespondenței. Formularele respective au fost expediate prin poștă la 10 milioane de oameni, fiecare fiind rugat să completeze chestionarul și să-l expedieze înapoi.

Modul în care au fost aleși cei 10 milioane de alegători a fost și sursa erorii foarte mari în rezultatul prezis. Astfel, au fost selecționați:

1°. fiecare al treilea alegător înregistrat din Chicago;

2°. fiecare alegător înregistrat din Pennsylvania și Allentown.

Numele au fost extrase la întîmplare astfel:

1°. din listele membrilor diferitelor cluburi;

2°. din cartea de telefon;

3°. din listele diferitelor asociații;

4°. din membrii diferitelor consilii orășenești etc.

Ceea ce a încercat să facă „The Literary Digest" a fost alcătuirea așa-numitei „selecții naționale". Se constată imediat că selecția utilizată pentru analiză nu a fost reprezentativă în sensul că, printre cei selectați, majoritatea erau persoane cu venituri anuale ridicate, în timp ce majoritatea alegătorilor, luată ca un întreg, nu prezenta deloc această caracteristică.

În acest mod, predicția a fost, în mod serios, afectată de tendința acestui grup particular de alegători, de a vota în general în favoarea forțelor conservatoare ale societății americane din acel timp, reprezentată de A. D. Landon.

Pe de altă parte, din nou metoda corespondenței s-a dovedit destul de ineficientă deoarece numai *o pătrime* din corespondenți au returnat chestionarele completate.

Problema „selecției naționale" nu este deci numai o problemă de statistică, ci și una de sociologie. După cum s-a văzut, nu are importanță prea mare numărul elementelor din selecție, dacă acesta nu este astfel ales încît structura socială — sub toate aspectele ei — să fie proporțional reprezentată în această selecție.

Bineînțeles că selecția națională nu are un rol limitat la cadrul unor alegeri electorale. Ea poate fi folosită cu succes în studiile de marketing (prospectarea pieței), de demografie etc.

Azi, sondajele de opinie publică în diferite domenii au devenit o practică uzuală în multe țări din lume. O faimă meritată și-a cucerit-o, printre alte institute similare de acest tip, Institutul pentru sondarea opiniei publice George Gallup din S.U.A. Totuși, și acest celebru institut a suferit un eșec — în anul 1948 — tot cu ocazia estimării unor alegeri prezidențiale. Eroarea nu a fost atît de mare ca în cazul „The Literary Digest", învățămintele trecutului fiind folosite de către Gallup.

Ceea ce a contribuit la insuccesul sondajului din 1948 a fost considerarea în selecția națională a prea multe persoane cu studii superioare în raport cu cele cu studii medii și elementare și, de asemenea, încheierea anchetei cu un timp prea larg înaintea alegerilor propriu-zise, timp în care unele persoane și-au schimbat atitudinea, s-au retras de pe listele electorale etc., ceea ce, evident, a influențat votul real față de analiza statistică prealabilă.

După cum se observă, problema obținerii unei selecții naționale reprezentative este o problemă suficient de complicată și depinde, evident, de condițiile fiecărei țări în parte, de problemele concrete ce se pun în fiecare țară. De exemplu, alta va fi structura socială a selecției naționale într-o țară în curs de dezvoltare față de aceea într-o țară puternic industrializată.

Alcătuirea selecției naționale trebuie făcută de către o echipă interdisciplinară de specialiști, deoarece numai statisticianul nu poate cuprinde multiplele aspecte ale problemei.

Există, evident, și alte fațete ale aplicării metodelor statistice, și, în general, a matematicii în sociologie.

Trebuie să facem observația că modelarea fenomenelor sociale este destul de dificilă tocmai datorită multitudinii de factori ce determină sau influențează aceste fenomene. Toate aceste modele, pentru a fi cît mai aproape de realitate, trebuie să aibă un caracter probabilistico-statistic ca și natura însăși a fenomenelor socio-demografice.

Să ne gîndim, de pildă, la studiul mobilității forței de muncă sau la studiul privind mobilitatea atitudinilor față de o anumită problemă, de exemplu, simpatia față de o anumită echipă de fotbal. Astfel, dacă avem o colectivitate compusă din N indivizi, să presupunem că ne interesează atitudinea acesteia față de o echipă de fotbal, echipă pe care s-o numim — pentru a nu supăra pe nimeni — X.

În mod normal, atitudinea (simpatia) fiecărui „microbist" rămîne neschimbată o oarecare perioadă de timp. Dar, după această perioadă, din motive în general subiective, atitudinea se poate schimba. Ceea ce interesează este estimarea numărului de microbiști care, la un moment dat, vor avea o atitudine de simpatizanți ai echipei X.

Problema este, în esență, o problemă de teoria probabilităților și statistică matematică.

Exemplul nostru este, poate, „de agrement", însă nu e greu de ghicit că pot fi găsite exemple similare dar de importanță reală.

Aici cercetările sînt deschise și există multe aspecte care trebuie rezolvate — atît teoretice cît și practice — astfel încît modelele statistice privind „mobilitatea atitudinilor" să fie efectiv aplicabile[1].

[1] Vezi: Postelnicu, V., *Modele pentru studiul mobilității interramuri*. În: „Revista de Statistică", nr. 4, 1970, pp. 68—72.

2.2. STATISTICA ȘI PRODUCȚIA INDUSTRIALĂ

> *„Posibilitatea existenței unei științe este determinată nu prin ce, ci prin cum; nu prin obiect, ci prin metodă. De aceea, noi credem că tehnica, deci și industria, poate fi în mod sigur obiectul unei științe — și anume a unei științe foarte importante."*
>
> M. Korach [1]

Multă vreme, producția industrială a fost considerată ca opusă științei, iar disputa asupra rolului istoric al științei, pe de o parte, și al tehnicii, pe de altă parte, continuă încă și astăzi.

Engels a sesizat de mult interdependența dintre știință și tehnică și a intuit în mod corect rolul producției industriale în dezvoltarea societății omenești.

Într-o scrisoare devenită celebră el spunea:

„Dacă ... tehnica depinde în cea mai mare parte de stadiul în care se află știința, apoi într-o și mai mare măsură depinde aceasta de stadiul și necesitățile tehnicii.

Dacă societatea are o necesitate tehnică, ea împinge știința înainte mai mult decît zece universități". (s.n.) [2]

Geniala remarcă a lui Engels își găsește o ilustrare strălucită în ceea ce privește statistica.

Chiar dacă problemele legate de populație, de anchete sociale, de biologie sau genetică au contribuit la dezvoltarea statisticii în mod de netăgăduit, cea căreia statistica îi este profund îndatorată este producția industrială.

Acest lucru nu este valabil numai pentru statistică. Tot Engels, în *Dialectica naturii* făcea observația:

„Pînă acum se făcea paradă numai cu ceea ce producția îi datorează științei: *știința însă îi este infinit mai îndatorată producției*". (s.n.)

---

[1] Korach, M., *Știința industriei*. În: „Știința despre știință", Editura politică, București, 1968, p. 197.

[2] Marx, K., Engels, Fr., *Opere alese*, vol. II, E.S.P.L.P., București, 1955, p. 552.

Să urmărim cum aceste idei ale lui Engels își găsesc reflectarea în evoluția istorică a statisticii.

Necesitatea tehnică — și socială — a societății, la un moment dat al dezvoltării sale, o constituia găsirea acelor mijloace prin care să se ajungă să se producă mai mult, lucru care s-a realizat odată cu introducerea mașinismului, deci odată cu apariția mecanizării proceselor de producție.

Știința a profitat de pe urma acestei necesități prin dezvoltarea mecanicii, fizicii, chimiei — domenii impulsionate direct de revoluția industrială.

Din primele decenii ale secolului nostru, și cu deosebire în zilele noastre, problema a devenit mai complexă în sensul că pe lîngă cantitate interesează, cu deosebire, aspectul legat de calitate. Nu de calitatea obținută oricum, ci de calitatea obținută în mod economic.

Aceasta este „necesitatea tehnică" a societății industriale a zilelor noastre. Nu trebuie să ne gîndim neapărat la un „aparat" sau un „mecanism" anume pentru fabricarea calității. Calitatea, într-adevăr, se produce, se fabrică dar nu numai cu ajutorul mașinilor — oricît de complicate și moderne ar fi ele. Alături de mașină trebuie să conlucreze metodologii specifice, concepții și, bineînțeles, omul — factor ce nu-și pierde importanța nici în era celei mai complexe automatizări.

Vă întrebați, poate, ce caută statistica în acest context? Ei bine, statistica a progresat și a putut să-și aducă aportul la dezvoltarea însăși a producției tocmai datorită imperativului calității. Nu trebuie să se tragă concluzia că numai datorită problemelor legate de calitatea producției, știința statisticii a urmat un mers ascendent. Aceste probleme însă au scos în evidență rolul major pe care trebuie să-l joace metodele statistice în ansamblul general al metodelor de conducere a economiei și societății.

Este util să reamintim aici care a fost evoluția inițială a *Controlului statistic al calității* (SQC — în abreviere internațională), căci, în esență, despre el este vorba.

Iată cum s-au petrecut lucrurile: în jurul anilor 1920, o mare corporație americană de aparataj telefonic, și anume „Western Electric", se găsea într-o situație dificilă datorită faptului că o mare parte a produselor sale erau respinse la control ca necorespunzătoare din punct de vedere calitativ.

Pe vremea aceea, compania avea un așa-numit „Engineering Department" — care în 1925 a primit numele de Bell Telephone Laboratories — care grupa diverși specialiști în probleme de producție și care constituia și grupul de cercetare al companiei respective.

În 1924, tînărului cercetător Walter A. Shewhart (1891—1967), care lucra în acest laborator, i s-a cerut să găsească o soluție de reducere a acestei proporții mari de obiecte respinse la control, ca necorespunzătoare.

În urma studiilor întreprinse, Shewhart a prezentat un raport (în același an) în care a pus bazele așa-numitului *control statistic de fabricație* (mai corect: în timpul fabricației), care, practic, a însemnat nașterea unei noi ramuri a statisticii și anume controlul statistic al calității proceselor tehnologice.

Principiile care au stat la baza elaborării noii metodologii au fost expuse de Shewhart în 1931, în lucrarea *Controlul economic al calității produselor industriale* și mai tîrziu, în 1939, în colaborare cu W. E. Deming el și-a expus concepțiile despre statistică în lucrarea *Metoda statistică din punctul de vedere al controlului calității.*

Înainte de a vedea ce aduce nou Shewhart și cum intervine statistica, este interesant să observați următorul fapt: lui Shewhart nu i s-a cerut „vă rugăm să elaborați controlul statistic al procesului tehnologic folosind pentru aceasta cutare metodă statistică", ci, de fapt, rezolvarea unei probleme de economie: cum să facem să reducem pierderile.

Pentru că, în esență, conducerea oricărei întreprinderi industriale este interesată să producă mai mult, mai bine și mai ieftin. Din această cauză, matematicianul care intră într-o unitate productivă va fi la început dezamăgit deoarece va întîlni acolo doar mașini

ce lucrează, oameni ce le mînuiesc și nicidecum, atîrnînd din utilaj în utilaj, probleme de programare liniară, de teoria probabilităților sau statistică matematică, de mecanică sau de teoria elasticității. Nici specialistul în problema concretă și nici conducătorul întreprinderii nu-i vor spune „să vedeți, avem aici de-a face cu următoarea problemă de statistică și trebuie s-o rezolvăm cu metoda X pe care n-o cunoaștem suficient".

Ceea ce se cere, este o problemă la obiect: cum să se realizeze un beneficiu mai mare din vînzarea produselor, cum să se folosească la maximum capacitățile existente, cum să fie respinse de beneficiar cît mai puține produse etc.

Rolul de a descoperi natura — statistică sau nu — a unei astfel de probleme, precum și de a formula problema în termeni statistici sau în alți termeni matematici, revine matematicianului — în cazul de față celui ce se ocupă de statistica matematică.

Să vedem acum în ce a constat noutatea ideilor lui Shewhart. În primul rînd, a pornit de la premisa că un obiect, odată produs, nu mai poate fi schimbat calitativ decît printr-o cheltuială în plus: sau trebuie reparat („rectificat"), sau înlocuit cu unul de același tip, de calitate corespunzătoare.

Controlul calității *produselor,* ca atare, are deci un rol mai mult sau mai puțin pasiv, el reprezentînd, într-un anume sens, o oglindă a procesului de producție respectiv.

Ceea ce trebuie, de fapt, controlat și menținut sub control este însuși procesul de fabricație, deoarece este evident că un proces tehnologic necorespunzător nu va putea produce obiecte de calitatea cerută.

Acest control nu trebuie să fie oricum: el trebuie să fie operativ și economic. Operativ înseamnă rapid și simplu, iar economic înseamnă, în primul rînd, să coste puțin.

Metoda lui Shewhart — contrar tuturor celor ce încă se miră că ceva simplu[1] poate asigura astfel de

[1] Dar neapărat bine aplicat.

cerinţe — reprezintă un mod de a contribui la îmbunătăţirea calităţii produselor în mod economic.

Ideea „curat naivă" — cum ar fi spus unicul Caragiale — că în situaţia actuală cineva îşi poate permite să instaleze la fiecare strung un laser sau o altă minune electronică pentru a controla bucată cu bucată orice banală piesă cilindrică, ar conduce la preţuri de cost astronomice.

Desigur, astfel se pune problema la unicate sau la produse de serie foarte mică cum ar fi, de pildă, rachetele spaţiale. Controlul calităţii rachetelor Soiuz sau Apollo sau orice alt nume nu se va face în nici un caz „statistic". Acolo orice componentă va fi, poate, controlată „electronic", piesă cu piesă, pentru că în acest caz nu se permite nici o proporţie, oricît de mică, de defecte. Aceasta nu înseamnă că rachetele n-au voie să se defecteze; intervine aici problema siguranţei în funcţionare — o altă faţetă a calităţii.

Dar pentru producţia de mare serie — pentru un banal papuc de casă, de exemplu — dacă vom instala aparatură electronică pentru controlul cusăturii, fiecărui papuc în parte, atunci mai mult ca sigur că la preţul fiecăruia ar trebui să mai adăugăm cel puţin un 0.

Am analizat problema controlului înainte de a vedea efectiv în ce constă metodologia shewhartiană, pentru că există, în special printre statisticieni, părerea că într-o serie de ţări foarte avansate din punct de vedere industrial, cum ar fi S.U.A. sau Japonia (unde controlul statistic are o amploare deosebită, succesele înregistrate fiind chiar mai mari decît în ţara de origine a lui Shewhart), controlul se face automat [1], renunţîndu-se la metodele statistice.

Să revenim deci la Shewhart; el a pornit de la ideea că orice proces tehnologic este caracterizat, în esenţă, de doi parametri — să le zicem statistici — de bază:

[1] Cu privire la controlul automat, articolul omului de ştiinţă ceh J. Hrabák, *Reglage Statistique Automatique*. În: „Revue de Statistique Appliquée", vol. XII, nr. 3, 1964, pp. 105—110) lămureşte foarte clar problema. Se pare însă că lucrarea nu e prea cunoscută de cei ce propovăduiesc anacronismul metodologiei shewhartiene.

1°. *Centrarea procesului* — reprezentată printr-o valoare medie a valorilor caracteristicii de calitate luată în studiu.

Valoarea medie este, după cum ştim, un indicator al tendinţei centrale a unei serii statistice.

2°. *Precizia procesului* — reprezentată prin dispersia, prin împrăştierea valorilor aceleiaşi caracteristici.

Dispersia, şi în aceeaşi măsură „abaterea standard" (rădăcina pătrată din dispersie — vezi manualul), este un indicator al concentrării (sau al împrăştierii) unui set de măsurători experimentale.

Orice dereglare a unui proces tehnologic se produce, în fond, la nivelul unuia (în cazul cel mai rău al ambilor) dintre aceşti factori.

Sînt cunoscute expresii — ce azi au devenit uzuale — ca: „maşina sau utilajul cutare este descentrat", „nu este precisă", „are o precizie foarte redusă" etc.

Cei ce lucrează într-o întreprindere productivă întîlnesc la tot pasul problema importantă a reglării utilajelor, reglare care se face tocmai cu scopul de a aduce procesul tehnologic la centrarea şi precizia cerută. Problema nu constă în reglajul în sine, căci acesta se învaţă ca orice altă meserie şi, mai devreme sau mai tîrziu, ajunge o muncă de rutină: problema este *cînd* şi *cît de des* să se efectueze reglajele. Pentru că acest lucru nu este indiferent: orice reglaj înseamnă oprirea utilajului deci oprirea producţiei. Din nou chestiunile economice.

Dar, metodologia shewhartiană şi tehnicile statistice auxiliare ne ajută şi în această problemă: să producem calitatea dar în mod economic. Instrumentul practic de acţiune este aşa-numita „fişă de control de tip Shewhart". Ideea de bază este simplă — ca de altfel toate ideile mari: folosirea pas cu pas a informaţiei conţinută în măsurătorile experimentale în vederea descoperirii cauzelor care provoacă abaterea procesului tehnologic de la centrarea şi precizia dorită — mai precis, dintre limitele acceptabile ale acestor parametri fundamentali.

Revoluţionar în gîndirea shewhartiană, nu este folosirea globală a informaţiei asupra procesului, ci utili-

zarea succesivă a unei informații parțiale în scopul acționării imediate asupra procesului în vederea corectării lui. În acest mod, fără să o precizeze, Shewhart a dat *„primul exemplu de cibernetică statistică"* — după cum a subliniat statisticianul indian P. C. Mahalanobis, într-o lucrare intitulată sugestiv *Controlul calității pentru dezvoltarea economică*[1].

Sîntem datori să lămurim mai concret lucrurile: iată, să presupunem că pe un strung se prelucrează niște piese cilindrice și că ceea ce ne interesează este caracteristica „diametru", adică lungimea acestui diametru. Să considerăm că se produc un număr oarecare de astfel de piese și că, la un moment dat, se constată că piesele nu mai corespund cu cerințele fixate adică au diametrul mai mare.

Ce este de făcut? Este clar că trebuie făcut ceva, dar nu se știe ce anume și cînd să se acționeze.

Desigur, fără a fi mari specialiști în prelucrarea metalelor, ne dăm seama că o ipoteză ar fi tocirea sculei așchietoare (este evident că un cuțit tocit va furniza un diametru mai puțin precis decît unul ascuțit). Pe de altă parte, s-ar putea tot atît de bine ca materialul prelucrat să fie de vină, adică pe banda de prelucrare să fi intervenit un material mai dur pe care cuțitul, deși încă utilizabil, să nu poată să-l prelucreze la cota dorită.

În timp ce se fac presupuneri asupra cauzei, utilajul produce în continuare și e de la sine înțeles că singur nu se va corecta. Chiar dacă luăm o hotărîre bună, bazată pe o gîndire tehnică la obiect, este mai mult ca sigur că această hotărîre o vom lua, sau mai tîrziu decît trebuie, sau, în unele cazuri, mai devreme. Deci, implicit, nu vom produce în mod economic: sau vom pierde timp, sau vom irosi material.

Ce ne propune metodologia shewhartiană în această situație?

---

[1] Mahalanobis, P. C., *Quality control for economic growth*. În: „The Indian Journal of Statistics", Ser. B, vol. 29, nr. 3—4, 1967, pp. 191—200.

La început, dacă procesul tehnologic este bolnav, trebuie să-i găsim diagnosticul adecvat.

Aceasta o realizăm prin intermediul fișelor de control, care se construiesc pentru „tendința centrală" (medie, $\overline{X}$) și pentru „precizie" (abaterea standard, estimată prin amplitudine, R).

Iată cum: să presupunem că în cazul nostru, din oră în oră măsurăm diametrele a cîte 4 sau 5 piese cilindrice produse. (Să considerăm că luăm cîte 5; motivul îl vom vedea imediat). Formăm deci grupuri de cîte 5 măsurători:

1	2	...	16
x_{11}	x_{21}		$x_{16.1}$
x_{12}	x_{22}		$x_{16.2}$
x_{13}	x_{23}		$x_{16.3}$
x_{14}	x_{24}		$x_{16.4}$
x_{15}	x_{25}		$x_{16.5}$

Am ales 16 grupuri fiindcă am considerat ca exemplu două schimburi de lucru.

Pentru fiecare din aceste grupuri calculăm media:

$$\overline{X}_1 = \frac{x_{11} + x_{12} + \dots + x_{15}}{5}, \dots, \overline{X}_{16} = \frac{x_{16.1} + x_{16.2} + \dots + x_{16.5}}{5} \qquad (47)$$

precum și amplitudinea:

$$R_1 = \max\{x_{11}, \dots, x_{15}\} - \min\{x_{11}, \dots, x_{15}\}, \qquad (48)$$

$$\vdots$$

$$R_{16} = \max\{x_{16.1}, \dots, x_{16.5}\} - \min\{x_{16.1}, \dots, x_{16.5}\}. \qquad (49)$$

Tendința centrală se estimează deci prin media generală:

$$\overline{\overline{X}} = \frac{1}{16}(\overline{X}_1 + \dots + \overline{X}_{16}). \qquad (50)$$

Este necesar acum să studiem amplitudinea. Astfel, dacă avem un șir de date exprimentale:

$$x_1, x_2, \ldots, x_n$$

adică, în fond, niște numere, este limpede că unul dintre ele va fi cel mai mic, iar altul va fi cel mai mare[1]. Adică:

$$\max \{x_1, x_2, \ldots, x_n\} = x_{max}, \qquad (51)$$

$$\min \{x_1, x_2, \ldots, x_n\} = x_{min}. \qquad (52)$$

Diferența dintre x_{max} și x_{min} este tocmai amplitudinea. Se notează cu R (de la englezescul „range").

Deci,

$$R = x_{max} - x_{min}. \qquad (53)$$

Amplitudinea joacă un rol important în statistică. După cum este știut, dispersia în cazul repartiției normale se estimează cu ajutorul expresiei:

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2, \qquad (54)$$

unde x_i sînt datele de observație, $\bar{x}$ media lor, iar n numărul lor.

Abaterea standard (σ) se estimează prin extragerea rădăcinii pătrate din expresia lui σ^2, adică:

$$\hat{\sigma} = \left[\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2\right]^{1/2}. \qquad (55)$$

Deci, $\hat{\sigma}$ se exprimă, de fapt, în aceleași unități de măsură ca și media. Dar, R, amplitudinea, se exprimă

[1] Desigur, un formalist ar spune că putem avea $x_1 = x_2 = \ldots = x_n$ și atunci $x_{max} = x_{min} = x_i$, pentru orice $i = 1, \ldots, n$. În practică această situație se va întîmpla extrem de rar. Și chiar dacă se va întîmpla, nu schimbă cu nimic raționamentul.

în aceleași unități ca și media fiindcă și ea este o expresie „liniară".

Cu ajutorul amplitudinii se poate estima abaterea standard. Acest lucru se demonstrează riguros în statistica matematică. Noi ne vom menține aici doar la observația intuitivă a unei unități de măsură comune între abaterea standard și amplitudine.

Expresia lui $\hat{\sigma}$ calculat cu ajutorul lui R, nu este greu de imaginat:

$$\hat{\sigma}_R = \frac{R}{dn}, \qquad (56)$$

unde d_n este un coeficient calculat în funcție de volumul selecției adică în funcție de numărul datelor de observație.

Se remarcă faptul că dacă vom calcula σ cu ajutorul lui $\hat{\sigma}_R$ va fi mult mai ușor decît dacă vom întrebuința radicalul din dispersie.

Statisticienii au constatat însă că dacă mărim volumul selecției, atunci „precizia" lui $\hat{\sigma}_R$ față de adevărata valoare a lui σ scade, rezultate foarte bune obținîndu-se pentru valori mici ale lui n (în general, $n = 4, \ldots, 8$).

Iată deci că nu întîmplător Shewhart a ales, pentru calcule, grupuri de cîte 5 date.

Pe de altă parte, observați un lucru nu mai puțin lipsit de importanță: cum calculați foarte repede media a 5 numere? Le adunăm, bineînțeles, dar apoi, în loc să le împărțim cu 5, înmulțim suma cu 2 și împărțim cu 10. Lucrul poate să vă pară ciudat, dar gîndiți-vă la rapiditatea procedeului și la faptul că datorită simplității sale el poate fi aplicat chiar de lucrătorul care efectuează operațiile de prelucrare ale unui anumit produs.

Acum urmează construirea fișelor de control propriu-zise: pentru medie — tendința centrală — și pentru precizie, estimată prin $\bar{R}$ (practic, prin media celor

16 amplitudini — în cazul nostru). Din această cauză, fişele de control de tip Shewhart mai sînt cunoscute şi sub numele de „fişele de control ($\bar{X}$, R)".

Se stabilesc limitele pentru tendinţa centrală şi pentru precizie, între care, cu o probabilitate de 0,99 sau 99%, vor fi cuprinse majoritatea valorilor medii ale caracteristicii de calitate luată în considerare.

Fişele se construiesc împreună după modelul de mai jos:

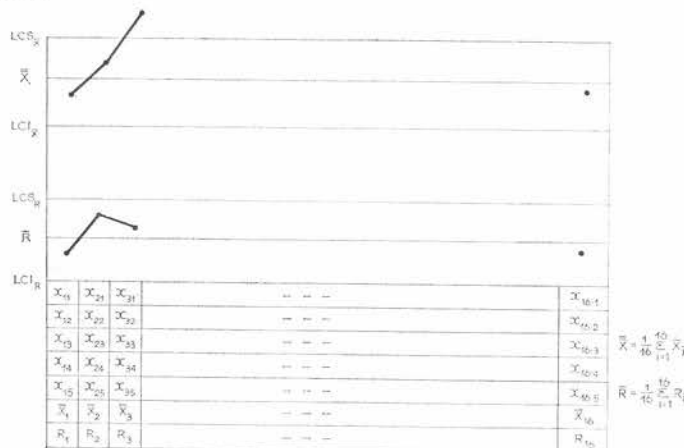


Fig. 11 — Un exemplu de construire a fişelor de tip Shewhart

Cum se calculează limitele „de control" este o problemă mai complicată care depăşeşte nivelul acestei lucrări. Dar, important este faptul că *acest calcul este standardizat*, adică în orice manual de control statistic al calităţii sau în normele interne ale fiecărei întreprinderi sînt prevăzute aceste calcule, astfel încît nu există nici un impediment. Pentru *acţiune* nu trebuie să ştim demonstraţii, ci să aplicăm corect ceea ce au demonstrat alţii, să înţelegem esenţa şi utilitatea acestor lucruri şi, bineînţeles, ceea ce e mai important, să cunoaştem bine procesul tehnologic.

În tabelul de mai jos se arată modul de construire al fişelor de control:

Metoda	Fişa pentru tendinţa centrală ($\bar{X}$)	Fişa pentru precizie (R)
Tendinţa centrală şi precizia estimate prin $\bar{\bar{X}}$ şi $\bar{R}$	Linia centrală $= \bar{\bar{X}}$ $LCS_{\bar{x}} = \bar{\bar{X}} + A_2R$ $LCI_{\bar{x}} = \bar{\bar{X}} - A_2R$	Linia centrală $= \bar{R}$ $LCS_R = D_4\bar{R}$ $LCI_R = D_3\bar{R}$

Tabelul 13: **Calculul fişelor de control de tip Shewhart**

Coeficienţii A_2, D_3 şi D_4 se extrag din tabele statistice speciale. În cazul nostru, în care am format grupuri de observaţii de volum 5 avem:

$$A_2 = 0{,}58, \quad D_3 = 0, \quad D_3 = 2{,}11. \tag{57}$$

Să mai observăm că limitele de control pentru $\bar{X}$ sînt (faţă de $\bar{\bar{X}}$) simetrice, pe cînd cele pentru R nu mai sînt simetrice faţă de $\bar{R}$.

Acest lucru se explică prin următorul fapt: dacă avem o serie statistică $x_1, x_2, ..., x_n$ care provine dintr-o populaţie normală de medie μ şi dispersie σ^2, atunci media $x = n^{n-1}\Sigma x_i$ urmează tot o lege normală dar de parametri μ şi σ^2/n. Adică, în fond, aceeaşi lege simetrică, dar de parametri diferiţi.

Cu amplitudinea, lucrurile stau altfel: chiar dacă seria iniţială provine dintr-o populaţie normală, repartiţia lui:

$$R = x_{max} - x_{min} \tag{58}$$

nu mai este normală.

Din cele de mai sus, aţi observat că fişele de control de tip Shewhart sînt calculate în ipoteza repartiţiei normale. Oricine poate însă obiecta: bine, dar nu este obligatoriu ca toate caracteristicile de calitate ale produselor să urmeze legea Gauss-Laplace (normală).

Este adevărat; dar aici intervine o teoremă fundamentală a statisticii matematice:

„Dacă o serie statistică $x_1, x_2, ..., x_n$ provine dintr-o populaţie oarecare, atunci media ei $\bar{X} = n^{-1}\Sigma x_i$ este o variabilă aleatoare care ascultă aproximativ de legea normală şi acest lucru se întîmplă chiar dacă volumul seriei statistice este mic".

Acest lucru esenţial a fost folosit de către Shewhart aşa că valabilitatea aplicării fişelor de control este universală, chiar dacă la început nu cunoaşte ce lege statistică urmează caracteristica de calitate ce o investigăm.

Să vedem acum, în fine, cum acţionează fişele de control. Iată, să presupunem deci că s-au produs 5 piese cărora le-am măsurat diametrul, am calculat media $\bar{X}$ şi am fixat pe grafic punctele respective. Punctele se află între limite, deci se poate produce în continuare.

Din nou acelaşi lucru. După a treia oră constatăm că punctul corespunzător pe fişa $\bar{X}$ se află în afara limitei de control superioare. Abrevierile *LCS* şi *LCI* înseamnă „limita de control superioară" şi respectiv „limita de control inferioară" (în engleză: *UCL* şi *LCL*).

Din acest moment, prelucrarea pieselor se opreşte şi se caută cauza care ar fi putut produce defecţiunea.

Precizia procesului rămînînd între limitele de control, putem bănui, sau că scula aşchietoare a suferit o uzură prematură, sau că în materia primă ceva s-a schimbat. Este acum rolul inginerului, al maistrului, al muncitorului să intervină cu cunoştinţele lui specifice, cu practica pe care o are pentru a depista cauza respectivă. Statistica devine acum „vioara a doua". După cum s-a mai spus de cîteva ori pînă acum, gîndirea tehnică nu poate fi înlocuită nici de cea mai înaltă matematică.

Desigur, în practică intervin şi probleme mult mai complicate decît prelucrarea unei piese. Fişele de control au însă o aplicabilitate universală: ele sînt un exemplu foarte sugestiv de modul în care statistica este, într-adevăr, un instrument practic de acţiune.

2.3. STATISTICA ÎN FIZICĂ ŞI CHIMIE

„Metodele statistice au pătruns în fizica clasică cu prilejul studiului proceselor moleculare, cînd datorită lucrărilor lui J. Maxwell, L. Boltzmann, R. Clausius, J. Gibbs şi ale altor savanţi a fost creat un nou capitol al ei — mecanica statistică.

Metodele statistice joacă un rol uriaş şi în fizica de astăzi. Statistica este o parte inalienabilă a tuturor teoriilor cuantice, care formează un capitol conducător al fizicii teoretice de astăzi" [1].

Din cîte am văzut pînă acum, nu este greu de dedus că statistica va juca un rol important în fizică şi chimie fiindcă aici avem, prin excelenţă, de-a face cu experimente, cu procese de măsurare care, prin natura lor, furnizează date statistice ce necesită o prelucrare în vederea obţinerii unei informaţii.

În multe probleme de fizică sau chimie, ne interesează stabilirea unor relaţii funcţionale între diferitele variabile ce intervin în experiment. Acest lucru se face tot cu ajutorul statisticii matematice.

În alte situaţii, chiar dacă fizica sau chimia ne spune ce anume relaţie există între factorii consideraţi, într-un caz concret, corespunzător unei anumite condiţii experimentale, va trebui să determinăm parametrii ce intervin în legea, în relaţia funcţională pe care o cunoaştem din fizica sau chimia teoretică.

Iată un exemplu pentru prima situaţie:

I. Se studiază căderea unei paraşute. Interesează legătura între viteza de cădere v (în metri pe secundă)

[1] (Sub redacţia K u z n e ţ o v, I. V. şi colaboratori), *Problema cauzalităţii în fizica modernă*, Edit. ştiinţifică, Bucureşti, 1963, p. 227.

și presiunea exercitată pe suprafața parașutei, p (în newtoni pe metrul pătrat).

Experimentul furnizează, de exemplu, următoarele date:

Nr. crt.	Viteza de cădere (m/s)	Presiunea pe suprafață (N/m²)
1	2,40	1381,8
2	3,50	2753,8
3	5,00	5507,6
4	6,89	11025
5	10,00	22050

Tabelul 13: Rezultatul experimentului privind lansarea unei parașute

Ce trebuie să facă experimentatorul?

Mai întîi, el va reprezenta valorile celor două variabile v și p într-un sistem ortogonal de axe pentru a vedea care este natura legăturii.

Iată, în fig. 12, acest grafic:

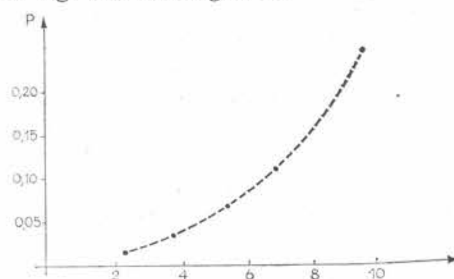


Fig. 12 — Reprezentarea grafică a datelor privind studiul căderii parașutei

Nu este greu de văzut că deși datele sînt puține, forma curbei $p = f(v)$ este parabolică. Așadar, se va alege ca legătură funcțională o parabolă de tipul:

$$p = \theta_0 + \theta_1 v^2, \tag{59}$$

Prin urmare, acum trebuie determinați parametrii θ_0 și θ_1, problemă care, în general, a mai fost discutată.

Pentru al doilea caz, considerăm următorul exemplu:

II. Din studiul chimiei teoretice se știe că numărul de ioni rămași într-un gaz, după acțiunea unui agent ionizant, este dat de formula:

$$N = \frac{N_0}{1 + N_0 \alpha t}, \tag{60}$$

unde N_0 este numărul de ioni prezenți în gaz, t — timpul, iar α o constantă specifică numită „coeficient de recombinare".

Să presupunem că efectuăm un experiment pentru determinarea coeficientului de recombinare a unui anumit gaz supus acțiunii ionizante a radiațiilor X.

La diverse valori ale lui t, obținem variații ale valorilor lui N.

Problema constă deci nu în stabilirea legăturii dintre N și t, ci în estimarea parametrilor legăturii care în cazul nostru sînt N_0 și α[1].

Vom aplica deci metoda celor mai mici pătrate, dar înainte vom face un mic artificiu pentru a ușura calculele. Scriem (60) astfel:

$$N^{-1} = \frac{1 + N_0 \alpha t}{N_0} = N_0^{-1} + \alpha t; \tag{61}$$

și acum, dacă notăm:

$$N^{-1} = Y, \quad N_0^{-1} = a, \quad \alpha = b,$$

atunci:

$$Y = a + b \cdot t \tag{62}$$

și deci putem aplica nestingheriți metoda ajustării liniare descrisă în manual.

[1] V o d ă, V., *Aplicații ale analizei de regresie în chimie*. În: „Revista de Chimie", vol. 25, nr. 9, 1974, pp. 719—723.

Nu trebuie să ne construim imaginea că metoda celor mai mici pătrate este atît de simplă precum pare în cazul liniar, sau în alte cazuri ce se pot reduce la cel liniar prin diferite transformări.

Să ne gîndim că în fizică sau în chimie există şi procese foarte complicate, care nu pot fi descrise decît prin dependenţe complicate de tipul:

$$Y = a_1 f_1(x) + a_2 f_2(x) + ... + a_n f_n(x), \quad (63)$$

unde $f_1(x), ..., f_n(x)$ pot avea forme oarecare ce depind de parametrii ce trebuie estimaţi.

Rezolvarea unei astfel de probleme — care poartă numele de problema generalizată a celor mai mici pătrate — nu mai este chiar atît de uşoară.

O formă particulară a lui (63) a fost rezolvată în 1943 de către Deming, cel despre care am mai vorbit.

În anul 1972, doi ingineri americani[1] au reuşit să stabilească un program pe calculator pentru rezolvarea acestei probleme, iar în anul 1974, Iacob M. Dan, student la Institutul Politehnic din Bucureşti, Facultatea de automatică, a reuşit să perfecţioneze programul celor doi americani şi să-l adapteze pentru rezolvarea „elegantă" a oricărei probleme de „teoria regresiei", cum numesc statisticienii acest capitol al dependenţelor funcţionale între variabilele aleatoare.

Ceea ce am prezentat pînă aici n-a fost decît o ilustrare a unor metode statistice aplicate în fizică şi chimie, metode care se pot aplica foarte bine în demografie, în industrie etc. Şi, ca să respectăm istoria, această „teorie a regresiei", asupra căreia am insistat atît, a fost la început aplicată în biometrie de matematicianul englez F. Galton în 1886, în studiul legăturii dintre înălţimea unui grup de indivizi şi cea a descendenţilor lor. Galton a constatat o tendinţă la generaţia următoare de a „regresa" către valoarea medie (înălţimea medie) şi, în acest mod, termenul de „regresie" a intrat în vocabularul statistic, menţinîndu-se pînă astăzi, fără însă vreo referire la sensul iniţial.

[1] Powell, D. R., Mac Donald, J. R. În: „The Computer Journal", vol. 15, nr. 2, 1972.

Fizica şi chimia sînt legate însă de statistică nu numai prin tehnici de calcul, ci şi prin mod de gîndire, prin modul de abordare al unor probleme fundamentale. Pentru a arăta acest lucru, avem nevoie de o mică incursiune în teoria cinetică a gazelor şi în termodinamică. Desigur, în liceu se studiază mai complet cele ce vom prezenta noi aici — ca teorie fizico-chimică.

Pornind deci de la teoria atomistă, ştim că materia este alcătuită din atomi liberi sau grupuri de atomi care formează molecule.

Dacă considerăm o masă de aer cald şi o masă de aer rece vom constata experimental, conform teoriei cinetice, că între ele există o diferenţă destul de mare, cu toate că structura lor chimică poate fi absolut identică. Diferenţa constă în faptul că moleculele masei de aer cald se mişcă dezordonat cu viteze foarte mari.

Un fizician[1] compara moleculele cu un roi de albine care a fost deranjat şi care se agită încoace şi încolo într-o mişcare neregulată. Dar, în vreme ce albinele dintr-un roi se pot feri una de alta, ca să nu se lovească, orice moleculă se deplasează conform legilor binecunoscute ale mecanicii: rectiliniu şi cu viteză uniformă. În traiectoria sa, o moleculă se loveşte, sau de o altă moleculă, sau de peretele recipientului în care se află masa respectivă de aer.

Iată deci — deşi oarecum ascuns — primul element probabilistico-statistic al acestei situaţii: datorită numărului mare de molecule, probabilitatea ca o moleculă să se ciocnească de o alta oarecare nu este nulă. Am putea spune, mai sugestiv, că tocmai datorită acestui număr mare, probabilitatea ca să nu se ciocnească (deci evenimentul contrar) este foarte mică.

Toate aceste ciocniri fac ca traiectoriile parcurse de molecule să fie foarte complicate, tocmai datorită schimbărilor dese de direcţie şi viteză.

Desigur, dacă s-ar reuşi să se localizeze exact poziţia şi deplasarea *tuturor* moleculelor, atunci s-ar putea imediat evalua ce molecule se vor lovi şi cum va arăta

[1] Peierls, R. E., *Legile naturii*, Editura ştiinţifică, Bucureşti, 1963, p. 113.

deplasarea lor după ciocnire. Imensul număr de molecule face însă ca ideea să nu fie deloc practică și aceasta chiar în ipoteza fericită că s-ar dispune de aparatura necesară efectuării unor astfel de observații. Deoarece numărul de molecule este enorm, nu ne vom mai concentra atenția asupra unei singure molecule.

Explicarea comportării fizice a unui gaz se face deci studiind comportamentul *în medie* al unui număr suficient de mare de molecule.

Din acest moment apare statistica: nu prin simple tehnici de calcul, ci prin gîndire. Folosirea gîndirii statistice este aici o necesitate obiectivă impusă de însăși realitatea fizică a fenomenelor considerate.

Apariția noii discipline, numită azi „fizica statistică" (inițial „mecanica statistică"), nu mai putea întîrzia, tocmai datorită faptului că fenomenele specifice comportării gazelor nu se mai puteau explica altfel decît cu ajutorul statisticii.

Astfel, în 1859, J. C. Maxwell a prezentat o comunicare la Asociația Britanică din Aberdeen în care, pentru prima oară, a aplicat statistica matematică la problemele mecanice ridicate de mișcarea moleculelor. În 1860 apare lucrarea sa fundamentală, *Illustrations of the dynamical theory of gases* (Scientific Papers, vol. 1, pp. 377—410), în care stabilește legea de repartiție a vitezelor moleculelor, care-i poartă numele, și care a generat ulterior cercetări fructuoase ce au desăvîrșit teoria cinetică a gazelor, lămurind pe deplin toate aspectele ce încă nu se puteau explica prin metodele deterministe.

Vorbind despre lucrările lui Maxwell, fizicianul Max von Laue spunea în *Istoria fizicii* [1].

„Cu aceasta, teoria cinetică a gazelor era constituită în trăsăturile ei fundametale; ...

Aceste trăsături fundamentale își au rădăcina în mecanica newtoniană. Și totuși, *odată cu teoria cinetică a gazelor a pătruns în fizică un punct de vedere nou: considerentele probabilistico-statistice*" (s.n.).

[1] L a u e, M. v o n, *Istoria fizicii*, Edit. științifică, București, 1963, p. 139.

Să vedem acum în ce au constat ideile lui Maxwell.

Se știe că într-un gaz oarecare, moleculele se mișcă liber, iar viteza de deplasare este diferită pentru fiecare moleculă. Pentru a putea aplica metodele statistice, ar trebui estimat numărul de molecule care au vitezele cuprinse între două limite bine stabilite. Acest lucru se poate determina practic, folosind o instalație experimentală numită „aparatul lui Lammert".

Gazului, constînd dintr-un ansamblu de molecule — deci dintr-o populație statistică — i se atașează o funcție de densitate $f(v)$, care reprezintă densitatea de probabilitate a vitezelor moleculelor, unde v este viteza absolută a acestora, iar $f(v)$ numărul moleculelor cu viteză v.

Maxwell a dovedit experimental că forma densității $f(v)$ este:

$$f(v) = B\,e^{-h^2v^2} \cdot v^2, \qquad (64)$$

unde B și h sînt constante pozitive ce depind numai de temperatură și de natura fizico-chimică a gazului.

Desigur, conform cu cele ce știm despre o densitate de probabilitate, trebuie să avem $f(v) \geqslant 0$ și $\int_0^\infty f(v)\,dv = 1$.

Prima relație este, evident, îndeplinită, iar cea de-a doua ne va furniza valoarea lui B și a lui h. Vom vedea însă că B și h nu se aleg numai din considerente strict matematice, pentru a avea satisfăcută egalitatea

$$\int_0^\infty f(v)\,dv = 1. \qquad (65)$$

Constantele B și h, avînd semnificație fizică, în primul rînd trebuie să o oglindească pe aceasta. La determinarea lui B și h a contribuit, în mod esențial, termodinamica, ramură a fizicii, deosebit de interesantă, care, de-a lungul vremii, a suscitat nenumărate dispute filozofice. Nu este scopul nostru să considerăm astfel de detalii; în orice manual de fizică vom găsi date aceste valori:

$$B = 4\pi\left(\frac{m}{2\pi kT}\right)^{3/2} \quad \text{și} \quad h = \left(\frac{m}{2kT}\right)^{1/2}, \qquad (66)$$

unde m este masa moleculei, T — temperatura măsurată în grade Kelvin, iar K = *constanta a lui Boltzmann*.

Graficul unei densități de tip Maxwell arată ca în fig. 13.

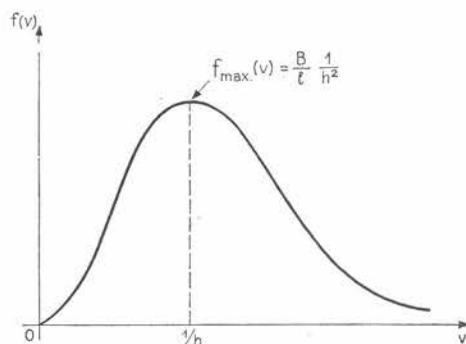


Fig. 13 — Densitatea de tip Maxwell

Să observăm că:

$$f(o) = 0, \lim_{v \to \infty} f(v) = 0, \quad (67)$$

$$f'(v) = 2Bv\, e^{-h^2v^2}(1 - h^2v^2)$$

și deci:

$$v_{max} = \frac{1}{h} \Rightarrow f_{max}(v) = \frac{B}{e} \cdot \frac{1}{h^2}. \quad (68)$$

Din punct de vedere fizic, figura de mai sus ne arată cum se repartizează moleculele gazului în funcție de viteze. Majoritatea moleculelor se grupează în jurul vitezei $1/h$, deoarece numărul maxim de molecule este $f_{max}\left(\frac{1}{h}\right)$ și are valoarea:

$$f_{max}(v) = \frac{1}{e} \cdot \frac{\sqrt{8m}}{\sqrt{\pi k}} \cdot \frac{1}{\sqrt{T}} \quad (69)$$

ținînd cont de expresiile constantelor B și h.

Ceea ce se observă este faptul că *maximul descrește odată cu creșterea temperaturii.*

În figura 14 se prezintă un caz concret și anume cazul mercurului gazos la două temperaturi:

$$\begin{cases} T_1 = 273° \ K(0°C) \\ T_2 = 373° \ K(100°C) \end{cases}$$

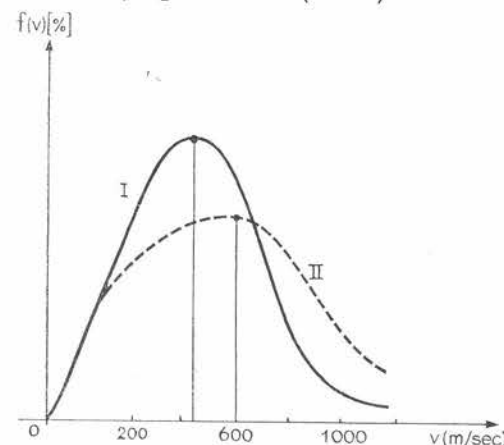


Fig. 14 — Densitățile de tip Maxwell în cazul mercurului gazos la două temperaturi diferite (I — la $t = 0°C$; II — la $t = 100°C$)

Ceea ce se observă din graficele experimentale, în cazul acestui gaz, este descreșterea lui $f_{max}(v)$, dar, de asemenea, și creșterea valorii v corespunzătoare, adică a vitezei moleculelor pentru care este obținut acest $f(v)$ maxim.

Pornind acum de la densitatea lui Maxwell, putem calcula valoarea medie a vitezei moleculelor într-un gaz oarecare.

Valoarea medie, conform definiției din manual, este:

$$E(v) = \int_0^\infty v \cdot f(v)\, dv. \quad (70)$$

Avem deci de calculat integrala definită:

$$E(v) = B \int_0^\infty v^3 \cdot e^{-h^2v^2}\, dv. \quad (71)$$

Integrarea se face folosind formula[1]:

$$\int x^n e^{-x^2} dx = -\frac{x^{n-1}e^{-x^2}}{2} + \frac{n-1}{2}\int x^{n-2}e^{-x^2} dx. \quad (72)$$

În cazul nostru $n = 3$ şi notînd $hv = x$, avem de calculat:

$$\int x e^{-x^2} dx \quad (73)$$

care este integrala din membrul II pentru $n = 3$. Dar,

$$\int x e^{-x^2} dx = -\frac{1}{2} e^{-x^2}. \quad (74)$$

În fine, mai considerăm acum şi limitele de integrare şi găsim:

$$E(v) = \frac{B}{2h^4} \quad (75)$$

sau, ţinînd seama de valoarea constantelor:

$$E(v) = \frac{2}{\sqrt{\pi}} \cdot \sqrt{\frac{2kT}{m}}. \quad (76)$$

Iată deci că, în medie, viteza moleculelor creşte odată cu creşterea temperaturii.

Mai putem calcula, de asemenea, valoarea medie a energiei cinetice a moleculelor:

După cum se ştie, formula energiei cinetice este $\frac{mv^2}{2}$. Astfel, avem iar de evaluat o integrală definită şi anume:

$$E\left[\frac{mv^2}{2}\right] = \frac{mB}{2}\int_0^{\infty} v^4 e^{-h^2v^2} dv. \quad (77)$$

Rezultă, în final:

$$E\left[\frac{mv^2}{2}\right] = \frac{3mB\sqrt{\pi}}{16h^5} = \frac{3}{2} kT, \quad (78)$$

formulă bine cunoscută din manualele de fizică.

[1] Smoleanski, M. L., *Tabele de integrale nedefinite*, Edit. tehnică, Bucureşti, 1972, p. 122.

Aşadar, energia cinetică medie a moleculelor unui gaz este proporţională cu temperatura absolută.

Iată una din legile importante ale fizicii dedusă pe calea raţionamentelor statistice.

Pe de altă parte, relaţia (78) arată că se poate determina o valoare medie a energiei cinetice prin simpla măsurare a temperaturii gazului, întrucît k este o constantă universală.

Ceea ce am prezentat pînă acum constituie doar un exemplu de gîndire statistică în fizică. Literatura menţionată în bibliografie cuprinde şi alte exemple de aplicare a statisticii în fizică şi chimie.

Capitolul III

3. EXEMPLE DE APLICAŢII CONCRETE ALE METODELOR STATISTICO-MATEMATICE

„... acolo unde matematica se termină, începe statistica."

Gopal Kanji [1]

3.1. EXEMPLE SAU APLICAŢII?

Acest capitol pare să înceapă cu un joc de cuvinte: exemple de aplicaţii. În acelaşi timp, ne punem întrebarea, care este deosebirea între un exemplu şi o aplicaţie, fiindcă, în mod uzual, această deosebire nu prea este sesizată — cel puţin în matematică.

Cînd se învaţă, de pildă, regula de derivare a unei funcţii de tipul $[u(x)]^{v(x)}$ şi se cere apoi să se deriveze $y = x^{\sin x}$, $x \in \left(0, \frac{\pi}{2}\right)$, atunci ceea ce facem constituie o ilustrare, un exemplu pentru formula sau metoda generală.

Aplicaţia este, după părerea noastră, un concept mai complex, şi poate datorită faptului că nu este întotdeauna bine înţeles, duce la regretabila prejudecată că „a face aplicaţii" ale matematicii — în cazul nostru ale statisticii matematice — înseamnă o muncă de rutină, lipsită de interes.

Vorbind despre W. A. Shewhart, profesorul W. E. Deming spunea:

„Shewhart a fost, cred, primul care a subliniat faptul că practica reclamă mai multă grijă, dacă nu şi mai

[1] Kanji, G. M., *The role of statistical laboratory in the teaching of statistics*. În: „Int. J. of Math. Educ. in Sci. and Techn.", vol. 5, nr. 1, 1974, p. 53.

multă adîncime a cunoaşterii în general, decît aceea cerută de cercetarea pură sau învăţămînt.

Filozofia, teoria cunoaşterii, problemele ce constituie baza teoretică pentru acţiune, toate acestea apar cu tărie atunci cînd cineva intră într-o activitate practică" [1].

A face o aplicaţie a metodelor statistico-matematice înseamnă, în primul rînd, existenţa unei probleme dintr-un domeniu specific: chimie, fizică, inginerie etc.

O problemă reală nu coincide aproape niciodată cu un exemplu sau cu o ilustrare a unui test de concordanţă; ceea ce vrem să subliniem este faptul că o problemă implică, de cele mai multe ori, un ansamblu de metode statistice pe care statisticianul trebuie să ştie cum şi în ce condiţii să le aplice.

Aceasta reclamă în mod necesar o cunoaştere a domeniului din care problema concretă trebuie să fie rezolvată; cunoaşterea, chiar dacă este sumară, trebuie însă să fie foarte clară, deoarece în bună măsură succesul aplicării statisticii depinde de modul în care s-au înţeles problemele specifice domeniului investigat.

Un alt aspect: o problemă nu va apare niciodată formulată în termeni statistico-matematici. Acest lucru l-am mai arătat aici, dar este interesant de remarcat că unul din reproşurile pe care statisticienii le aduceau şi încă le aduc celorlalţi specialişti este acela că nu-şi formulează problemele în termeni matematici, adică, în definitiv, că nu prezintă un model statistico-matematic al problemei respective.

Dacă acest lucru ar putea să se întîmple întotdeauna, atunci e foarte limpede că drumul de la model la rezolvarea problemei nu mai e deloc lung sau greu, fiindcă tratate de statistică se găsesc în toate bibliotecile şi deci statisticianul nu şi-ar mai găsi rostul.

Obligaţia de bază a statisticianului — spunea tot W. E. Deming — într-o lucrare intitulată sugestiv *Codul comportării profesionale* [2], constă în primul rînd

[1] Deming, W. E., *Boundaries of Statistical Inference*, op. cit., p. 652.

[2] Deming, W. E., *Code of Professional Conduct*. În: „Int. Stat. Rev., vol. 40, nr. 2, 1972, p. 215.

în „a-l ajuta pe beneficiar să-şi formuleze problema în termeni statistici ..."

A fi inginer, nu înseamnă neapărat că trebuie să fii expert în electronică, după cum a fi matematician nu înseamnă că eşti foarte priceput în geometrie proiectivă.

Faptul că azi, din ce în ce mai mulţi specialişti cunosc şi apreciază metodele statistico-matematice, nu scade cu nimic rolul statisticianului.

De multe ori, modelul formulat pentru rezolvarea unei probleme reclamă noi instrumente sau metode statistice sau adaptarea şi dezvoltarea celor deja cunoscute. Aceasta înseamnă deci cercetare fundamentală ce revine în mod evident în sarcina statisticianului. Unele aplicaţii necesită experimentări în care apar alte probleme statistice, cît şi probleme de ordin economic şi organizatoric.

Iată deci că „a face aplicaţii" nu înseamnă nicidecum rutină sau plictiseală. Într-o convorbire personală cu autorul, profesorul Gh. Mihoc spunea:

„În cercetarea statistică, ideile mari nu vin niciodată dacă stai la birou cu hîrtia şi creionul în faţă. Aplicaţiile sînt o nesecată sursă de noi probleme şi mari satisfacţii atunci cînd îţi aduci contribuţia la rezolvarea lor".

Nu rămîne decît să exemplificăm cele spuse.

Pentru că am promis, mai de mult, că vom reveni asupra unor probleme de teoria siguranţei în funcţionare a sistemelor, iată mai întîi un dialog (imaginar) între un inginer electronist (I) şi un statistician (S).

I: La noi în fabrică, după cum aţi văzut, se produc motoare de avioane. Partea pe care o produce secţia noastră este blocul electronic de comandă.

S: Da l-am văzut ... e o „jucărie" destul de mică.

I: Adevărat, dar să ştiţi că nu e deloc simplă şi, pe deasupra, nu e deloc ieftină.
Dar să continui: blocul acesta de comandă este, ca să zic aşa, inima avionului sau mîna dreaptă a comandantului de zbor. Prin el se recepţionează condiţiile tehnice ale zborului, el sesizează micile nereguli, el semnalizează situaţiile critice etc. De aceea ne interesează foarte mult testările de durabilitate ale blocului, adică, practic, durata medie de bună funcţionare a sa.
Dar nu numai atît: ne interesează mai cu seamă probabilitatea ca el să funcţioneze neîntrerupt cel puţin un număr fixat de ore, să zicem 10 000 de ore. Pe de altă parte, ne interesează care este numărul de ore de bună funcţionare dacă cerem, de pildă, ca această probabilitate să fie de 0,999.

S: Bănuiesc că aţi rezolvat într-un fel problema, fiindcă, după cîte ştiu, avioanele care zboară cu motoarele voastre zboară totuşi bine.

I: Evident, dar modul în care am rezolvat noi problema ne costă cam prea mult.

S: Încep să înţeleg; bănuiesc că doriţi să elaboraţi o testare a durabilităţii componentelor în care să supuneţi încercărilor mai puţine elemente.

I: Ceva de genul ăsta. Noi vrem însă ceva şi mai mult. Ştiţi doar cum se efectuează unele încercări de durabilitate pentru orice componente, nu numai electronice. Din lotul de producţie se aleg n elemente care se supun simultan la standul de probă, lăsîndu-le să funcţioneze pînă cînd sau se defectează toate, sau numai un număr r, stabilit în prealabil. Între timp, înregistrăm duratele de funcţionare ale elementelor defectate şi ne apucăm să facem calculele.

S: Vă întrerup. Trebuie neapărat să presupuneţi şi să verificaţi dacă aceste durate de funcţionare se supun unei anume legi de probabilitate?

I: Categoric! Altfel, bineînţeles că nu putem face nici un calcul. Noi am luat, ca de obicei, legea de repartiţie de tip Weibull:

$$F(x;\theta,k) = 1 - e^{-\theta x^k},\ x > 0;\ \theta,\ k > 0 \qquad (79)$$

şi am constatat că pentru cazul nostru, cel mai bine se potriveşte legea Rayleigh, adică pentru $k = 2$:

$$F(x;\theta) = 1 - e^{-\theta x^2},\ x > 0,\ \theta > 0. \qquad (80)$$

S: În cazul acesta nu cred că situația este atît de complicată. Puteți să vă fixați de la început un număr de elemente (r) pe care să le lăsați să se defecteze prin funcționare continuă, după care încetați testarea, adică exact ceea ce mi-ați zis mai înainte.

I: Într-un fel aveți dreptate, dar lucrurile nu sînt atît de simple.

În primul rînd, nu ne mai putem permite să lăsăm să se defecteze decît *un singur element.*

În al doilea rînd, ne-ar interesa, care să fie numărul de elemente pe care să le supunem testării, fiindcă orice element care funcționează consumă energie și se consumă și pe sine, deci costă bani.

Fiindcă, să știți, pe noi ne costă nu numai să producem, dar și să verificăm ceea ce producem. Or, dacă economisim la producție și risipim la experimentare, n-am realizat nimic!

S: E adevărat, nu mai este chiar așa de simplu cum mi se părea la început.

I: Cred totuși că ați înțeles problema.

S: În linii mari, da; trebuie să oprim testarea după defectarea primului element și care să fie numărul de elemente supuse încercărilor, astfel încît costul experimentării să fie cît mai mic.

I: M-ați înțeles exact, nu numai în linii mari!

S: Nu știu dacă putem rezolva pe loc; trebuie să ne mai întîlnim, să o reluăm încă o dată și să stabilim detaliile.

Nu știm dacă personajele noastre s-au reîntîlnit și dacă au rezolvat sau nu problema. Aceasta nu mai are importanță — după cum neesențiale sînt și avioanele din exemplul de mai sus.

Dacă discuția este imaginară, problema nu este așa în nici un caz.

În problema testării durabilității obiectelor scumpe, ideea de a pune costul experimentării drept criteriu fundamental este cît se poate de naturală.

Iată, să încercăm să reformulăm dialogul nostru în termeni statistici; pentru rezolvarea problemei va trebui să știm care este costul experimentului pe unitatea de timp (fie el C_1 lei pe oră) și costul supunerii unui element la testare (fie el C_2 lei).

Ca dată statistică de observație avem valoarea $t_{(1)}$, ce reprezintă durata de bună funcționare a elementului care s-a defectat și după care s-a oprit testarea.

Prin urmare, costul acestui procedeu de testare este:

$$C = C_1 \cdot t_{(1)} + C_2 \cdot n, \tag{81}$$

n fiind numărul de obiecte supuse încercării, pe care vrem să-l determinăm astfel încît C să fie minim.

Nu vom putea face direct acest lucru, deoarece se observă că (81) variază de la un experiment la altul datorită lui $t_{(1)}$. Așadar, este natural să minimizăm *costul mediu*, adică:

$$\bar{C} = C_1 E\,(t_{(1)}) + C_2 \cdot n, \tag{82}$$

E fiind simbolul pentru valoarea medie. Expresia $E(t_{(1)})$ va depinde de n (deoarece au fost încercate n elemente, după care încercarea s-a oprit la defectarea unuia dintre ele) și se poate calcula dacă se cunoaște repartiția timpilor de funcționare fără defecțiuni, adică populația statistică de bază.

Calculul nu este întotdeauna banal și în multe situații problema este încă deschisă din punct de vedere teoretic.

Noi am ales aici unul din cazurile practice — să le zicem simple — deși nici acest caz nu e chiar atît de simplu. L-am ales pentru a ilustra ideea prof. Gh. Mihoc, privitoare la bogata sursă de inspirație pe care o constituie practica — adevăratul stimulent al cercetărilor teoretice.

3.2. O APLICAȚIE ÎN CONTROLUL STATISTIC DE RECEPȚIE

„Controlul calității — un mod de a ajuta pe alții să se ajute singuri."

W. R. Pabst [1]

Cazul prezentat mai jos reprezintă o aplicație efectivă a metodelor statistico-matematice. Există multe lucrări interesante care descriu astfel de aplicații, efectuate de noi sau în alte părți ale lumii. Alegerea e dificilă. Vom rezolva totuși dilema oprindu-ne, cu totul subiectiv, asupra unei lucrări [2] elaborate în cadrul unei unități industriale din țară.

Uzinele de fibre chimice din Săvinești — care au jucat împreună cu alte întreprinderi un rol de pionierat în propagarea aplicării statisticii matematice în țara noastră — produc alături de celebra melană și așa-numitele „rețele cord", ce se folosesc la fabricarea anvelopelor, ca inserție.

Rețeaua este, de fapt, o țesătură din fire de relon și de bumbac. Elementul de rezistență îl constituie firele de relon, cele de bumbac avînd numai rolul de a menține paralelismul firelor de relon, cu scopul de a forma o suprafață plană.

Firele de relon, din care sînt executate rețelele, trebuie să îndeplinească anumite condiții tehnice — cum ar fi sarcina de rupere (rezistența la rupere), alungirea la rupere, contracția etc. Evident, aceste condiții sînt stabilite în STAS-uri, norme interne ș.a.m.d., în funcție de sortimentul fabricat. Tot aceste standarde reglementează condițiile de recepție precum și modul de efectuare a determinărilor privind caracteristicile tehnice de calitate.

---

[1] Pabst, W. R., *International aspects of Quality Control Activity*, in „Contributions to Statistics", Edited by C. R. Rao, Pergamon Press, Calcutta, 1963, p. 325.

[2] Iliescu, D. V., Vodă, V. G., *O metodă de reducere a numărului de probe în controlul statistic de recepție folosind analiza de regresie*. În: „Revista de Statistică", nr. 12, 1972, p. 13.

Determinările de laborator sînt executate pe fire extrase din rețea. În acest scop, la două role din lot se țese în plus un metru de probă. Acest metru de probă se împarte în două. O jumătate de metru de probă servește la determinările de laborator ale furnizorului, iar cealaltă jumătate, la determinările de laborator ale beneficiarului, pentru recepționarea produsului.

Aici avem, într-un fel, o situație specială: extragerea probelor pentru control nu se poate face cu totul întîmplător — din motive de ordin tehnologic — intervenind deci un oarecare grad de determinare. Mai concret, situația este următoarea: un lot livrat de rețele cord este în același timp și un lot de producție, el fiind format dintr-un anumit număr de role, de obicei 8—9.

Ideal ar fi să se controleze fiecare rolă în parte, pentru luarea deciziei de acceptare sau respingere a lotului.

Acest lucru însă nu este posibil deoarece:

1° ar conduce la un număr mult prea mare de încercări;

2°. din motive de economie extragerea probelor trebuie să fie făcută doar din prima și ultima rolă a lotului de producție pentru a nu perturba procesul tehnologic (din această cauză se și efectuează țeserea în plus a „metrului de probă" respectiv).

Să considerăm acum că dorim să controlăm un lot de rețele cord în ceea ce privește caracteristica de calitate „rezistența la rupere" — una din caracteristicile de bază.

Această caracteristică este, așa-zisă, cu „limitare inferioară", adică există o valoare minimă a rezistenței la rupere pe care rețelele trebuie s-o aibă pentru a corespunde calitativ.

Procedeul de control este următorul: se efectuează determinări la prima rolă (uzual 20 de măsurători individuale) și se face media lor. Fie $\overline{X}_1$ media măsurătorilor pentru prima rolă. În mod analog, fie $\overline{X}_2$ media măsurătorilor pentru ultima rolă.

Dacă L_i este limita inferioară de control, atunci condițiile de acceptare a lotului, privitor la caracteristica luată în studiu, sînt:

$$\overline{X}_1 \geqslant L_i \quad \text{și} \quad \overline{X}_2 \geqslant L_i. \tag{83}$$

Dacă aceste condiții sînt îndeplinite, se apreciază că și rolele intermediare sînt de calitate corespunzătoare. Presupunerea este rezonabilă, deoarece dacă între momentele producerii celor două role (prima și ultima a unui lot) intervine o cădere a procesului, aceasta nu trece neobservată. Pe de altă parte, un proces tehnologic nu poate trece, de obicei, de la o stare de cădere înapoi, la starea normală, fără intervenția operatorului (ca analogie, să ne gîndim, că un televizor care se defectează nu prea se repară singur !).

În fond, în ce constă problema pe care vrem s-o rezolvăm?

Presupunînd că s-a controlat prima rolă și că s-a obținut $\overline{X}_1 \geqslant L_i$, ne punem întrebarea: în ce condiții s-ar putea renunța la controlul ultimei role? Aceasta ar avea o implicație directă și anume reducerea numărului de determinări, adică, în fond, o economie, deoarece, după cum am mai subliniat și cu alte ocazii, nu numai a produce costă, dar și a controla această producție.

Avînd în vedere natura procesului tehnologic, în cazul analizat a apărut, în mod natural, ideea investigării dependenței între cele două medii $\overline{X}_1$ și $\overline{X}_2$ care, după cum știm acum (vezi paragraful 2.2.), sînt variabile aleatoare repartizate normal.

S-a presupus și s-a verificat ipoteza dependenței liniare între $\overline{X}_1$ și $\overline{X}_2$ așa cum este înfățișată în figura 15.

Ideea a constat în determinarea unei valori a variabilei $\overline{X}_1$ — notată $\overline{X}_{1\alpha}$ — pentru care dacă $\overline{X}_1 \geqslant \overline{X}_{1\alpha}$, nu mai este necesară operația de control pentru ultima rolă, considerîndu-se cu riscul ales α (vezi paragraful 2.1) că și $\overline{X}_2 \geqslant L_i$, decizia fiind de acceptare a lotului pe baza rezultatelor controlului primei role.

Din figura 15 se observă că pentru toți $\overline{X}_1 > \overline{X}_{1\alpha}$ rezultă cu probabilitatea $(1 - \alpha)$ că $\overline{X}_2 > L_i$ deci:

a°. dacă $\overline{X}_1 < L_i$, se refuză lotul fără a se mai controla cea de-a doua rolă;

b°. dacă $L_i \leqslant \overline{X}_1 \leqslant \overline{X}_{1\alpha}$, se controlează și a doua rolă, decizia de acceptare luîndu-se dacă $\overline{X}_2 \geqslant L_i$;

c°. dacă $\overline{X}_1 > \overline{X}_{1\alpha}$, se acceptă lotul fără a se mai controla cea de-a doua rolă.

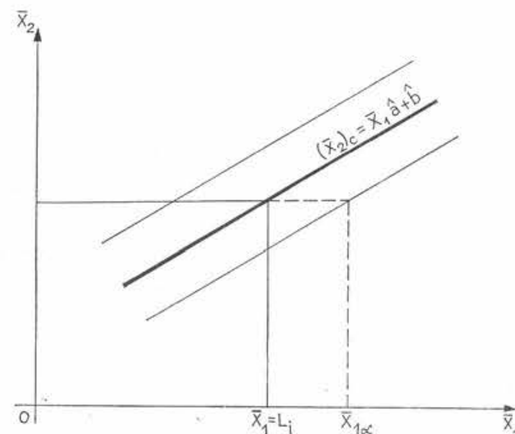


Fig. 15 — Dependența liniară și „zona de încredere" a liniei de regresie între cele două medii $\overline{X}_1$ și $\overline{X}_2$ reprezentînd rezistența la rupere a rețelelor cord

Iată deci cum cu ajutorul unei metode statistice se poate realiza o economie de timp, materiale și energie pentru întreprinderea respectivă. A nu face lucruri în plus — de exemplu, a nu controla a doua rolă atunci cînd nu este cazul — înseamnă, evident, o economie, iar în ultimă instanță utilitatea metodelor statisticomatematice se judecă tot prin prisma economică.

Acest aspect este valabil în orice domeniu al științei: nu este vorba de mercantilism, dar a crea, a face cercetare fără scopul de a aduce ceva nou și util în dezvoltarea producției materiale, este lipsit de sens.

În exemplul prezentat nu a fost explicată noțiunea de „zonă de încredere" care apare în fig. 15. Ea este oricum mai „tehnicistă" și nu este cazul să ne ocupăm de ea aici. Sperăm, bineînțeles, că ideea exemplului a fost înțeleasă și fără această noțiune.

3.3. O APLICAȚIE ÎN TEORIA SIGURANȚEI ÎN FUNCȚIONARE

„... istoria civilizației umane arată că fizionomia unei societăți este reconstruită pe baza produselor lăsate de aceasta."

D. Niculescu [1]

Dacă ne aducem aminte de cele discutate mai demult la teoria mortalității, afirmăm că instrumentul statistico-matematic utilizat acolo a constituit baza pentru dezvoltarea unei importante teorii — teoria siguranței în funcționare a sistemelor tehnice — care, în ultimele două decenii, a cunoscut o dezvoltare spectaculoasă. Această dezvoltare era naturală, ținînd cont de uriașul progres tehnic care a avut loc în ultimii 20 de ani.

Fiabilitatea (siguranța în funcționare) unui produs este unul din aspectele importante ale calității sale. Consumatorul este interesat ca produsul pe care-l cumpără să poată satisface scopurile pentru care acesta a fost creat. Un televizor trebuie „să meargă", nu numai să fie o „mobilă", iar acest „să meargă" trebuie să dureze cît mai mult.

Desigur, nimeni nu poate pretinde ca un televizor să „dureze" un sfert de secol. Probabilitatea ca el să funcționeze cel puțin un timp t_0 fixat, adică:

$$R(t) = \text{prob}\,\{T \geqslant t_0\}\,, \tag{84}$$

[1] N i c u l e s c u, D. *Calitatea — cerință obligatorie a planului de stat în anul 1974.* In: „Calitatea Producției și Metrologie", vol. IV, nr. 4, 1974, p. 197.

trebuie să fie cît mai mare. Aceasta este, în fond, fiabilitatea sistemului tehnic respectiv. Întreprinderile în care există un compartiment de control și fiabilitate a produselor care aplică în mod curent metodele statistice, fixează pentru produsele sale o perioadă t_0 în funcție de o serie întreagă de factori — calitate, costuri etc. — pe care o putem numi „perioadă de garanție" în care probabilitatea de „nefuncționare" adică $1 - R(t)$ este foarte mică. Din cîte știm pînă acum, $1 - R(t)$ este chiar funcția de repartiție a timpului de funcționare fără defecțiuni (există încă multe discuții pe plan mondial în ceea ce privește „perioada de garanție" și modul ei de definire; în mod uzual, prin ea se înțelege perioada de timp în care producătorul asigură eventuale reparații sau înlocuiri fără obligații bănești din partea beneficiarului. Această perioadă variază, în funcție de produsul respectiv: 6 luni, 1 an sau 2 ani. Nu intrăm aici în amănunte).

Dacă din punct de vedere matematic noțiunea de fiabilitate este simplă — vezi formula (84) — aspectele tehnice și economice ale acestei noțiuni nu sînt de fel simple.

Problema asigurării și menținerii siguranței în funcționare a produselor sau sistemelor tehnice are diferite aspecte care includ faza inițială de proiectare, controlul calității în timpul procesului de fabricație, controlul de recepție, încercările pe teren, modificările aduse la diferite etape ale proiectării sau fabricației. La acestea se mai poate adăuga problema optimizării costului afectat pentru asigurarea fiabilității sistemelor sau produselor.

Trebuie să mai observăm un lucru și anume: un sistem tehnic poate funcționa satisfăcător în anumite condiții și nu în altele, iar performanțele sale, satisfăcătoare pentru un anume scop, nu implică neapărat performanțe satisfăcătoare pentru alt scop. De exemplu, un tub electronic poate fi satisfăcător pentru un aparat TV, dar poate fi complet inutilizabil într-un sistem de ghidare a rachetelor.

În acest sens, siguranța de funcționare a unui produs — sau fiabilitatea acestuia — este definită ca fiind

probabilitatea ca el să funcţioneze cel puţin o perioadă de timp specificată, în condiţii de mediu cunoscute.

Cine a avut curiozitatea să demonteze un aparat de radio, un ceas sau orice alt aparat casnic la îndemînă, a remarcat, desigur, că aceste sisteme tehnice au componentele aranjate „în serie" sau „în paralel" sau, cel mai obişnuit, în combinaţii ale ambelor tipuri.

Un sistem de tip „serie" este un sistem în care componentele sale sînt legate astfel încît întregul sistem se defectează, dacă oricare din componentele sale se defectează. Iată, să considerăm sistemul format dintr-un buton de contact, un fir electric şi un clopoţel.



Fig. 16 — Un sistem tip „serie"

Presupunînd că totul este în regulă, dacă apăsăm pe buton, curentul electric trece prin fir şi acţionează clopoţelul (desigur, acest proces este mai complicat, dar pentru ceea ce vrem să ilustrăm, schema aceasta simplificatoare este suficientă). În momentul în care una dintre cele trei componente (buton, fir, clopoţel) se defectează, întreg sistemul este inoperant.

Să considerăm acum, mai general, un sistem format din n componente independente legate în serie:



Fig. 17 — Forma generală a unui sistem tip „serie"

În aceste condiţii, probabilitatea ca el să funcţioneze este dată de regula de înmulţire a probabilităţilor evenimentelor independente (vezi manualul de liceu) şi anume:

$$R_s = \prod_{i=1}^{n} R_i \qquad (85)$$

unde R_s este siguranţa sistemului, iar R_i este siguranţa componentei de rang i. Formula (85) pune în evidenţă deosebit de pregnant efectul creşterii complexităţii sistemului tip „serie" asupra siguranţei sale în funcţionare.

Exemplu: Dacă vom considera un sistem serie cu 4 componente, fiecare avînd siguranţa 0,95, atunci siguranţa sistemului este:

$$R_s = (0{,}95)^4 = 0{,}81.$$

Crescînd complexitatea sistemului la 8 componente de acelaşi tip, atunci siguranţa sa va fi:

$$R_s = (0{,}95)^8 = 0{,}66,$$

deci o scădere considerabilă a siguranţei sale în funcţionare.

Iată, prin urmare, că mărirea complexităţii unui astfel de sistem, prin adăugarea unor noi componente inseriate, nu duce la mărirea siguranţei sale în funcţionare.

Mărirea siguranţei în funcţionare se poate realiza printr-o altfel de aşezare a componentelor în sistem — atunci cînd acest lucru este posibil — şi anume prin aranjarea „în paralel".

Dacă un sistem este alcătuit din n componente independente legate în paralel, el se va defecta dacă toate cele n componente se vor defecta. Astfel, dacă $F_i = 1 - R_i$ este funcţia de „nesiguranţă" a componentei i, atunci putem aplica regula de înmulţire a probabilităţilor evenimentelor independente pentru a obţine funcţia de nesiguranţă a întregului sistem, şi anume:

$$F_p = \prod_{i=1}^{n} F_i. \qquad (86)$$

Pentru a deduce siguranţa sistemului, nu avem decît să scriem

$$R_p = 1 - F_p = 1 - \prod_{i=1}^{n} F_i = 1 - \prod_{i=1}^{n} (1 - R_i). \qquad (87)$$

Evident, în cazul în care $R_i = R$ pentru orice i, atunci:

$R_s = R^n$ (pentru cazul serie);

$R_p = 1 - (1 - R)^n$ (pentru cazul paralel).

Cele două formule pot fi folosite pentru calculul siguranței în funcționare a sistemelor „combinate".

Exemplu: Să considerăm un sistem format din 10 componente $C_1, C_2, ..., C_{10}$, fiecare avînd siguranța specificată în figura de mai jos:

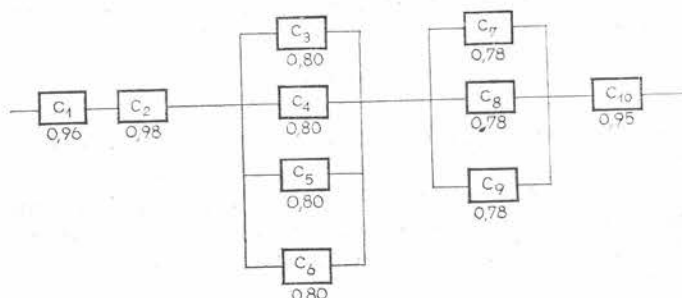


Fig. 18 — Exemplu de sistem combinat

În acest caz, subansamblul paralel $\{C_3, C_4, C_5, C_6\}$ poate fi înlocuit cu o componentă echivalentă K_1, avînd siguranța egală cu

$$1 - (1 - 0{,}80)^4 = 1 - (0{,}20)^4 = 1 - 0{,}002 = 0{,}998,$$

iar subansamblul $\{C_7, C_8, C_9\}$ poate fi înlocuit cu o componentă K_2 avînd siguranța egală cu

$$1 - (1 - 0{,}78)^3 = 1 - (0{,}22)^3 = 1 - 0{,}01 = 0{,}99.$$

Această înlocuire nu afectează, evident, siguranța generală a sistemului. Acum, sistemul nostru este alcătuit din componentele $C_1, C_2, K_1, K_2, C_{10}$ legate în serie și are deci siguranța

$$R_{\text{total}} = 0{,}96 \times 0{,}98 \times 0{,}998 \times 0{,}99 \times 0{,}95 = 0{,}883.$$

Se pune acum întrebarea: bine, calculăm siguranța în funcționare a unui sistem, fie el „serie", „paralel" sau

„combinat", dar de unde știm siguranța în funcționare a componentelor?

Este acum rîndul statisticii să intervină: din dialogul pe care l-ați citit în paragraful precedent, s-a putut constata că într-o anume fază a procesului de producție are loc o testare a durabilității componentelor sistemelor tehnice și chiar a sistemelor însăși.

Testarea durabilității componentelor și sistemelor este astăzi o disciplină independentă. Ea se numește în limbaj internațional „life-testing" („testarea duratei vieții") și poate fi considerată, prin natura metodelor folosite, ca o ramură a statisticii matematice.

Una din problemele esențiale în acest „life-testing" este determinarea repartiției timpului de funcționare fără defecțiuni a elementelor testate. Deci, din nou este vorba de stabilirea concordanței, doar că aici, spre deosebire de demografie, de exemplu, trebuie deseori să lucrăm cu date puține — cu „selecții de volum mic" în limbaj statistic, iar problemele statistice în acest caz sînt mult mai complicate decît cele așa-numite de „statistică clasică".

Un exemplu de acest gen am întîlnit chiar în dialogul statisticianului cu inginerul, din paragraful precedent. Acolo era vorba chiar de un caz extrem, cînd aveam la dispoziție doar o singură observație.

În capitolul de „life-testing", un loc aparte îl ocupă deci teoria repartițiilor statistice, luate drept model pentru timpul de funcționare fără defecțiuni. Teoretic, orice repartiție statistică pozitivă, adică orice variabilă aleatoare x cu densitatea $f(x)$ definită pentru $x > 0$ poate fi aleasă drept model. Totuși, din punct de vedere practic, numărul acestora nu este nelimitat. În alegerea unui model statistic pentru acest tip de probleme, se pleacă în mod natural de la conceptul de „rată a căderii" (intensitatea de defectare — vezi rata mortalității de la demografie):

$$\lambda(t) = \frac{f(t)}{1 - F(t)}, \tag{88}$$

unde $F(t)$ este funcția de repartiție asociată lui $f(t)$.

În practică, intervin fenomene variate în legătură cu procesul de defectare. Unele obiecte prezintă „căderi" rapide, de îndată ce sînt puse să funcţioneze. De exemplu, cuţite care se rup într-un proces de aşchiere, tuburi de radio sau TV care se ard imediat ce aparatul este pus în funcţiune; apoi, la majoritatea obiectelor avem de-a face cu fenomenul de uzură datorit unei folosinţe îndelungate etc. Forma funcţiei $\lambda(t)$ o cunoaştem de la teoria mortalităţii, numai că aici denumirile sînt altele.

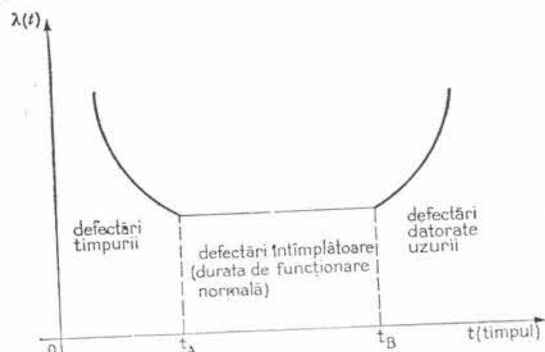


Fig. 19 — Forma tipică a ratei căderii (funcţia numită popular „cada de baie")

Ştim, tot de la teoria mortalităţii, care este legătura dintre funcţia de siguranţă (acolo o numeam funcţia de supravieţuire) şi rata căderii:

$$R(t) = e^{-\int_0^t \lambda(u)du}. \qquad (89)$$

Dacă vom presupune o rată a căderii constantă, adică exact de forma cuprinsă între t_A şi t_B din fig. 19, ceea ce înseamnă de fapt:

$\lambda(t) = \lambda = \text{constant} > 0$ (o dreaptă paralelă cu axa Ot),

atunci funcţia de siguranţă are forma simplă:

$$R(t) = e^{-\lambda t}, \quad t > 0, \quad \lambda > 0, \qquad (90)$$

iar funcţia de repartiţie, respectiv densitatea timpului de funcţionare fără defecţiuni, este:

$$F(t) = 1 - e^{-\lambda t}, \quad f(x) = \lambda e^{-\lambda t}, \quad t > 0, \quad \lambda > 0. \qquad (91)$$

Repartiţia de această formă este foarte bine cunoscută de către practicienii care lucrează în domeniul teoriei siguranţei şi se numeşte „repartiţia exponenţială". Ea are o îndelungată „carieră" în domeniul fiabilităţii, explicată în parte şi prin simplitate în ceea ce priveşte tratarea matematică. Nu întotdeauna însă fenomenele reale pot fi modelate prin modele simple. Dacă în unele situaţii repartiţia exponenţială este adecvată, în altele ea nu mai este deloc, iar analiza unui fenomen printr-un model neadecvat nu poate furniza o concluzie conformă cu realitatea. Este tocmai cazul componentelor care prezintă defectări premature, precum şi al fenomenului de uzură, în care rata căderii este descrescătoare, respectiv crescătoare.

În anul 1951, inginerul suedez Wallodi Weibull a scris un articol intitulat *O repartiţie cu o largă aplicabilitate*[1] în care a propus ca alternativă la modelul exponenţial, o repartiţie — ce azi îi poartă numele — care are densitatea de probabilitate:

$$T : f(t; \lambda, K) = \lambda K t^{K-1} e^{-\lambda t^K}, \quad t > 0; \quad \lambda, K > 0. \qquad (92)$$

Această repartiţie prezintă, faţă de repartiţia exponenţială, două avantaje importante; în primul rînd, rata căderii nu mai este constantă, ci funcţie de timp:

$$\lambda(t) = \lambda K t^{K-1}, \qquad (93)$$

fapt care este mult mai în concordanţă cu fenomenele reale, iar în al doilea rînd, prezenţa unui nou parametru — K — permite adaptarea mai elastică la diferite situaţii întîlnite în practică.

Pentru $K = 1$, după cum se observă din formula (92), se obţine chiar repartiţia exponenţială, iar pentru

[1] W e i b u l l, W., *A statistical distribution function of wide applicability*. În: „Journal of Applied Mechanics", nr. 18, 1951, pp. 293—297.

$K = 2$, repartiția Rayleigh, cea pe care inginerul din paragraful precedent o folosea în modelarea timpului de funcționare fără defecțiuni a blocurilor de comandă ale avioanelor, produse de secția pe care o conducea.

Iată cum arată $f(x; \lambda, K)$ pentru diferite valori ale lui K și λ fixat, să zicem $\lambda = 1$ pentru ușurința reprezentării:

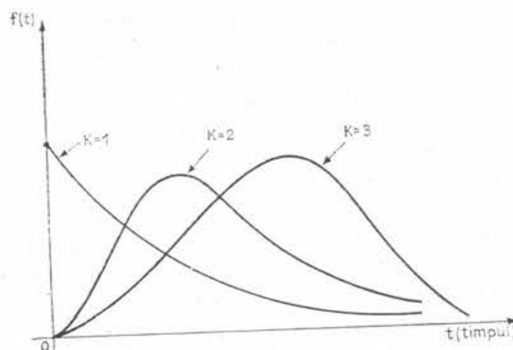


Fig. 20 — Densitatea de tip Weibull pentru $\lambda = 1$ și $k = 1, 2, 3$

Problema inițială, care se pune deci în fața celui care analizează fenomenele de fiabilitate, este aceeași ca și în orice alt domeniu de aplicare a metodelor statistice: stabilirea concordanței.

Sînt datele pe care le avem la dispoziție provenite dintr-o populație de tip Weibull, de exemplu? Desigur, putem scăpa foarte ușor de astfel de întrebări, presupunînd de la bun început că avem de-a face cu o lege exponențială, caz în care toate lucrurile sînt aparent corecte.

J. N. Berrettoni, un savant italian, printre primii care au folosit lucrările lui Weibull în industrie, spunea:

„Atenția noastră trebuie să fie îndreptată acum asupra realismului ipotezelor făcute în cadrul modelului statistic ales pentru analiză.

În teoria siguranței, repartiția exponențială este folosită extensiv și frecvent, fără a ține cont dacă presupunerea unei rate a căderii constantă este adecvată sau nu"[1].

Acest „realism al ipotezelor" este determinat, în primul rînd, de experiența practică, completat apoi de cunoașterea teoriei statistice. Chiar dacă, de multe ori, metodele efective de calcul nu depășesc cunoștințele ce se predau în licee, înțelegerea fenomenului și aplicarea corectă a teoriei statistice face parte din ceea ce noi numim *gîndirea statistică.*

Sîntem acum în măsură să ne dăm seama că acest proces de gîndire statistică este mult mai complicat decît demonstrarea unei teoreme în matematica pe care, în mod uzual, o numim pură.

Nimeni nu neagă frumusețea, ingeniozitatea, talentul care trebuie cheltuit într-o teoremă de geometrie plană, în celebrele locuri geometrice, de exemplu, sau în divizibilitatea numerelor. Dar faptul că trei ceviene într-un triunghi sînt sau nu concurente nu implică o răspundere deosebită. Dacă însă un produs oarecare este livrat cu indicația „probabilitatea ca el să funcționeze neîntrerupt 10 000 de ore este 0,98" cînd în realitate probabilitatea respectivă nu este decît 0,50, poate avea importante repercusiuni și grave răspunderi, atunci cînd produsul prin nefuncționarea sa conform cu specificațiile și pe perioada de timp considerată, pune în pericol viața celui ce-l utilizează.

Această situație poate apărea atunci cînd informația furnizată este extrasă pe baza unui model statistic neadecvat, depărtat simțitor de ceea ce se întîmplă în realitate.

Pentru a da un exemplu sugestiv asupra importanței alegerii corecte a unui model statistic în teoria siguranței, să considerăm cazul simplificat a două funcții de siguranță:

$$R_1(t) = e^{-t}, \qquad t \geqslant 0, \tag{94}$$

$$R_2(t) = 1 - t^2, \qquad 0 \leqslant t \leqslant 1. \tag{95}$$

[1] Berrettoni, J. N., *Foresteps to a reliable conclusion.* In: „Proc. Int. Symp. on Quality Control", JUSE, Tokyo, 1969, p. 620.

Ați recunoscut, desigur, în $R_1(t)$ funcția de siguranță a modelului exponențial $F(t;\ \lambda) = 1 - e^{-\lambda t}$, $t \geq 0$, pentru $\lambda = 1$, iar în $R_2(t)$ cea corespunzătoare pentru repartiția putere $F(t;\ \delta, b) = \left(\frac{t}{b}\right)^{\delta}$, $0 \leq t \leq b$, unde $b = 1$ și $\delta = 2$.

Despre această funcție s-a mai discutat în capitolul consacrat primelor aplicații ale statisticii.

Am ales $b = 1$ pentru ușurința reprezentării grafice; numărul 1 în sine, în acest caz, poate fi gîndit ca unitate de timp aleasă, de exemplu, 100, 1 000, 10 000 de ore.

Să privim acum graficele celor două funcții:

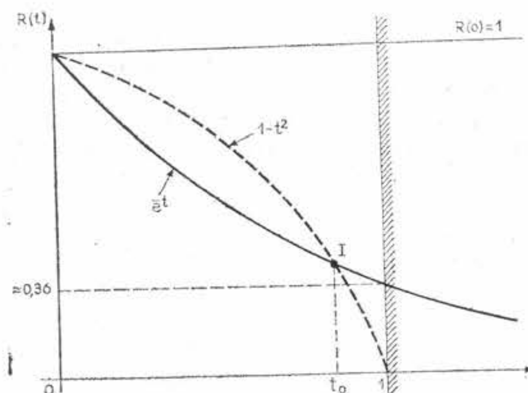


Fig. 21 — Două funcții de siguranță: e^{-t} și $1 - t^2$.

Reprezentarea grafică este foarte simplă. Cel puțin $1 - t^2$, $0 \leq t \leq 1$ este un banal trinom de gr. II cu maximul în (0, 1).

Proiecția pe axa Ot a punctului I de intersecție a lui $R_1(t)$ cu $R_2(t)$, notată de noi cu t_0, este tocmai soluția ecuației $e^{-t} = 1 - t^2$. Din figură, se observă imediat că:

1°. pentru valorile $t \in [0, t_0]$, $R_2(t) \geq R_1(t)$;

2°. pentru valorile $t \in (t_0, 1)$, $R_2(t) < R_1(t)$.

Ce semnificație practică ar avea acest fapt? În fond, $R(t)$ nu reprezintă altceva decît o probabilitate și anume probabilitatea ca anumite elemente, componente, produse etc. să funcționeze cel puțin un timp egal cu t.

Dacă ne-ar interesa în mod special numai valoarea t_0,[1] atunci, cum $R_1(t_0) = R_2(t)$ care după figură — presupunînd că am executat-o la scară — este aproximativ 0,40, evident nu are importanță ce model am fi ales: exponențial sau putere.

Să presupunem însă că în realitate modelul este putere, dar noi, conform tradiției — cum spunea Berrettoni — l-am ales pe cel exponențial. Este clar atunci că în mod nejustificat am fi indicat pentru orice $t \in [0, t_0]$ o probabilitate de bună funcționare mai mică decît în realitate, în timp ce pentru $t \in (t_0, 1)$ am fi indicat o probabilitate mai mare decît în realitate, lucru mai periculos în unele situații decît primul caz.

Iată cum, două reprezentări grafice ale unor funcții mai mult decît simple, pot pune în evidență unele implicații pe care nici un manual de analiză matematică nu le poate pune.

Să revenim acum la problema pe care ne-o punem referitor la repartiția Weibull: cum constatăm că datele de observație provin dintr-o populație de tip Weibull și ce facem în continuare. Din 1951 și pînă în prezent au apărut foarte multe lucrări consacrate acestei repartiții astfel că, după cum spunea tot un statistician, s-a manifestat o „euforie de tip Weibull". Există azi destule teste pentru verificarea concordanței în cazul Weibull, numeroase metode de estimare a parametrilor λ și K și multe alte probleme statistice legate de subiectul în cauză. Din păcate, toate acestea depășesc cu mult materia de liceu, necesitînd solide cunoștințe de statistică matematică. Totuși, repartiția Weibull nu este deplin pierdută pentru cauza noastră; dacă presupunem că datele ce le avem sînt într-adevăr dintr-o colectivitate de tip Weibull, atunci putem arăta cum se poate

[1] Sau valori într-o vecinătate suficient de mică a lui t_0, deoarece eroarea obținută este neglijabilă.

face estimarea parametrilor λ şi K într-un caz concret, numeric, folosind metoda celor mai mici pătrate pe care o cunoaştem acum destul de bine.

Să pornim de la următoarea problemă: Presupunem că în laboratorul de încercări al unei întreprinderi sosesc 100 de componente dintr-un lot de un anumit tip, a căror durată medie de funcţionare vrem s-o determinăm. Se aşază toate aceste elemente pe un banc de probă şi se pun să funcţioneze simultan. Nu vom putea aştepta ca toate să se defecteze şi apoi să facem media timpilor de funcţionare, din mai multe motive:

1°. ar dura prea mult în cazul în care componentele prin natura lor au o durabilitate mare;

2°. ar costa prea mult în cazul în care aceste componente ar fi relativ scumpe;

3°. n-am mai avea ce folosi ulterior în structura unor sisteme mai complexe sau efectiv nu s-ar mai putea vinde aceste componente ca atare.

Din aceste motive, întregul experiment se opreşte după defectarea unui număr dinainte fixat de componente (prin analiza prealabilă a costurilor şi a altor factori — vezi „dialogul S-I"), să zicem $r = 11$.

Să presupunem că s-au obţinut următoarele valori numerice:

i	1	2	3	4	5	6	7	8	9	10	11
$t_{(i)}$ (ore)	4	25	60	80	92	128	201	258	300	390	450

Tabelul 14: Timpii de funcţionare (în ore) a componentelor studiate

Aici i reprezintă rangul celei de-a i-a componente defectate, iar $t_{(i)}$ timpul de funcţionare.

Dacă adunăm valorile $t_{(i)}$, nu putem obţine o imagine reală a duratei medii a componentelor lotului, deoarece nu ştim legea de repartiţie statistică pe care o urmează timpii de funcţionare fără defecţiuni. Valoarea $(t_{(1)} + t_{(2)} + \ldots + t_{(11)})/11$ nu ne spune decît că durata medie a celor 11 componente $t_{(1)}, t_{(2)}, \ldots, t_{(11)}$ este „atît". Puterea statisticii se face simţită abia atunci cînd pe baza acestor 11 date putem „efectua o inferenţă" (adică o prelungire a rezultatelor gîndirii noastre) asupra întregului lot de produse şi aceasta cu un suficient de mare grad de rezonabilitate.

Cunoaşterea „comportamentului statistic" — adică, practic, a repartiţiei timpului de bună funcţionare — permite atît evaluarea duratei medii, a siguranţei în funcţionare cît şi a altor indicatori importanţi. Să presupunem că acest comportament este descris de o repartiţie de tip Weibull:

$$T : F(t; \lambda, K) = 1 - e^{-\lambda t^K}, \quad t > 0, \quad \lambda, \ K > 0, \tag{96}$$

ai cărei parametri λ şi K trebuie să-i estimăm din datele de observaţie. Iată cum vom proceda: scriem mai întîi (96) sub forma:

$$e^{\lambda t^K} = \frac{1}{1 - F(t; \lambda, K)}. \tag{97}$$

Logaritmăm acum în baza e şi obţinem:

$$\lambda t^K = \ln \frac{1}{1 - F(t; \lambda, K)}. \tag{98}$$

Logaritmăm din nou în aceeaşi bază:

$$\ln \lambda + K \ln t = \ln \ln \frac{1}{1 - F(t; \lambda, K)}.$$

Să notăm acum: $\ln \lambda = \hat{\alpha}_1$, $K = \hat{\alpha}_2$, $Y = \ln \ln \frac{1}{1 - F(t)}$.

În final se obţine:

$$Y = \hat{\alpha}_1 + \hat{\alpha}_2 \ln t. \tag{99}$$

Necunoscutele sînt $\hat{\alpha}_1$ şi $\hat{\alpha}_2$ pe care putem să le estimăm cu metoda celor mai mici pătrate, pe care am aplicat-o pînă acum în mai multe ocazii (vezi capitolul consacrat teoriei mortalităţii precum şi al aplicaţiilor în chimie).

Formăm deci suma:

$$S = \sum_{i=1}^{r} (Y_i - \hat{\alpha}_1 - \hat{\alpha}_2 \ln t_{(i)})^2 \qquad (100)$$

unde $Y_i = \ln \ln \frac{1}{1 - F(t_{(i)})}$, $t_{(i)}$, $i = 1, 2, ..., r$ (în cazul nostru $r = 11$) sînt valorile experimentale pe care le avem la dispoziție.

Luăm apoi derivatele $\frac{\partial S}{\partial \hat{\alpha}_1}$ și $\frac{\partial S}{\partial \hat{\alpha}_2}$, le egalăm cu zero și obținem următorul sistem de ecuații:

$$\begin{cases} r\hat{\alpha}_1 + \left(\sum_{i=1}^{r} \ln t_i\right)\hat{\alpha}_2 = \sum_{i=1}^{r} Y_i \\ \left(\sum_{i=1}^{r} \ln t_i\right)\hat{\alpha}_1 + \left(\sum_{i=1}^{r} \ln^2 t_i\right)\hat{\alpha}_2 = \sum_{i=1}^{r} Y_i \ln t_i. \end{cases} \qquad (101)$$

Bineînțeles, acest sistem este foarte simplu de rezolvat și regula lui Cramer sau, eventual, metoda reducerii, ne conduce în final la soluția:

$$\hat{\alpha}_1 = \frac{\left(\sum_{i=1}^{r} Y_i\right)\left(\sum_{i=1}^{r} \ln^2 t_i\right) - \left(\sum_{i=1}^{r} \ln t_i\right)\left(\sum_{i=1}^{r} Y_i \ln t_i\right)}{r\sum_{i=1}^{r} \ln^2 t_i - \left(\sum_{i=1}^{r} \ln t_i\right)^2}; \qquad (102)$$

$$\hat{\alpha}_2 = \frac{r\sum_{i=1}^{r} Y_i \ln t_i - \left(\sum_{i=1}^{r} Y_i\right)\left(\sum_{i=1}^{r} \ln t_i\right)}{r\sum_{i=1}^{r} \ln^2 t_i - \left(\sum_{i=1}^{r} \ln t_i\right)^2}. \qquad (103)$$

Așadar, estimațiile lui λ și K, adică de fapt estimațiile parametrilor repartiției Weibull sînt, prin acest procedeu:

$$\hat{\lambda} = e^{-\hat{\alpha}_1} \text{ și } \hat{K} = \hat{\alpha}_2. \qquad (104)$$

De acum, urmează calculele pentru deducerea efectivă a valorilor $\hat{\lambda}$ și $\hat{K}$. În acest scop, se întocmește tabelul 15.

i	$F(t_{(i)})$	$t_{(i)}$	$\ln t_{(i)}$	$\ln^2 t_{(i)}$	Y_i	$Y_i \ln t_{(i)}$
1	0,005	4	1,38629	1,92180	−5,29831	−7,34499
2	0,015	25	3,21887	10,36112	−4,19969	−13,51826
3	0,025	60	4,09434	16,76362	−3,68887	−15,10349
4	0,035	80	4,38202	19,20209	−3,35240	−14,69028
5	0,045	92	4,52178	20,44649	−3,10108	−14,02240
6	0,055	128	4,85203	23,54219	−2,90041	−14,07287
7	0,065	201	5,30262	28,11778	−2,73336	−14,49397
8	0,075	258	5,54907	30,79218	−2,56026	−14,20706
9	0,085	300	5,70345	32,52934	−2,44208	−13,92828
10	0,095	390	5,96357	35,56417	−2,30345	−13,73678
11	0,105	450	6,10702	37,29569	−2,20376	−13,45841
Σ	—	—	51,08106	256,53647	−34,78367	−1,457679

Tabelul 15: Calculul necesar estimării parametrilor legii Weibull în exemplul considerat

Să explicăm cum s-au calculat unele coloane ale acestui tabel; de exemplu, în coloana a VI-a apar valorile Y_i care sînt de forma:

$$\ln \ln \frac{1}{1 - u}. \qquad (105)$$

Se poate verifica destul de ușor că dacă $u > 0$, este suficient de mic, atunci poate fi făcută aproximarea:

$$\ln \ln \frac{1}{1 - u} \cong \ln u \qquad (106)$$

ceea ce, după cum se observă, ușurează foarte mult calculele. Să încercăm să „demonstrăm" această aproximație.

Presupunem că (106) este de fapt o egalitate; atunci putem scrie:

$$\ln \frac{1}{1-u} = u, \quad u > 0. \tag{107}$$

Derivăm ambii membri în raport cu u și avem:

$$\frac{1}{1-u} = 1, \tag{108}$$

egalitate valabilă numai dacă $u \to 0$. Dar se observă că dacă u este foarte mic, adică $0 < u \ll 1$, atunci eroarea survenită prin aproximația respectivă este foarte mică, deci putem utiliza (106).

În coloana a doua, apare $F(t_i)$, adică funcția de repartiție estimată în punctul t_i. Funcția de repartiție și estimația ei cu ajutorul datelor de observație este prezentată în manualul de liceu. Totuși, pentru a ilustra complet modul de întocmire al tabelului, reamintim formula:

$$\widehat{F(t_{(i)})} = \frac{i - \frac{1}{2}}{n} \quad i = 1, 2, \ldots, r, \tag{109}$$

cu ajutorul căreia s-a întocmit coloana a II-a.

În fine, putem acum calcula elementele ce intră în expresiile lui $\hat{\alpha}_1$ și $\hat{\alpha}_2$. Avem:

$$r \sum_{i=1}^{r} \ln^2 t_i = 11 \times 256{,}53647 = 2821{,}90117;$$

$$\left(\sum_{i=1}^{r} \ln t_i\right)^2 = (51{,}08106)^2 = 2609{,}27469;$$

$$m \sum_{i=1}^{r} \ln^2 t_i - \left(\sum_{i=1}^{m} \ln t_i\right)^2 = 212{,}62648;$$

$$\left(\sum_{i=1}^{r} Y_i\right)\left(\sum_{i=1}^{r} \ln^2 t_i\right) = -\,34{,}78367 \times 256{,}53647 =$$
$$= -\,8923{,}27992;$$

$$\left(\sum_{i=1}^{r} \ln t_i\right)\left(\sum_{i=1}^{r} Y_i \ln t_i\right) = 51{,}08106 \times$$
$$\times (1148{,}57679) = -\,7589{,}45993;$$

$$\left(\sum_{i=1}^{r} Y_i\right)\left(\sum_{i=1}^{r} \ln^2 t_i\right) - \left(\sum_{i=1}^{r} \ln t_i\right)\left(\sum_{i=1}^{r} Y_i \ln t_i\right) =$$
$$= -\,1333{,}81999.$$

Rezultă, deci:

$$\hat{\alpha}_1 = \frac{-\,1333{,}81999}{212{,}62648} = -\,6{,}27307.$$

Pentru calculul lui $\hat{\alpha}_2$ avem:

$$r \sum_{i=1}^{r} Y_i \ln t_i = 11 \times (-\,148{,}57679) = -\,1634{,}34469;$$

$$\left(\sum_{i=1}^{r} Y_i\right)\left(\sum_{i=1}^{r} \ln t_i\right) = -\,34{,}78367 \times 51{,}08106 =$$
$$= -\,1776{,}78673;$$

$$r \sum_{i=1}^{r} Y_i \ln t_i - \left(\sum_{i=1}^{r} Y_i\right)\left(\sum_{i=1}^{r} \ln t_i\right) = 142{,}44204;$$

$$\hat{\alpha}_2 = \frac{142{,}44204}{212{,}62648} = 0{,}669915.$$

Așadar:

$$\hat{K} \cong 0{,}67, \text{ iar } \hat{\lambda} \cong e^{-6{,}27} \cong 0{,}0018.$$

Deci, forma funcției de repartiție este următoarea:

$$F(t) = 1 - e^{-0{,}0018 \cdot t^{0{,}67}}, \quad t > 0. \tag{110}$$

Cineva ar putea pune acum întrebarea: de ce atîta strădanie pentru a obține forma (110)? Calculele noastre nu au fost zadarnice. Formula de mai sus este tocmai oglinda comportamentului statistic al duratelor de bună funcționare a componentelor studiate.

Formula (110) ne permite — printre altele — calculul duratei medii de funcționare, element important pentru beneficiarul care va utiliza mai departe aceste componente — fie în structuri mai complexe, fie ca atare.

Durata medie de funcționare, de fapt valoarea medie a variabilei t (repartizată Weibull în cazul nostru), este:

$$E(t) = \int_0^\infty t \cdot f(t;\lambda, k)\mathrm{d}t \text{ sau echivalent:}$$

$$E(t) = \int_0^\infty t \cdot \mathrm{d}F(t;\lambda, k), \qquad (111)$$

adică, în fond, momentul de ordinul I al lui t. În cazul nostru, un calcul ceva mai lung ne dă în final:

$$E(t) \cong 11000 \text{ ore.}$$

Se constată astfel că, în realitate, durata medie de funcționare a componentelor respective este de aproximativ 11000 ore și nu de aproximativ 181 ore (o valoare aproape de 70 de ori mai mică!) cum ar rezulta din $\sum_{i=1}^{11} t_i/11$.

Iată deci cum cunoașterea comportamentului statistic ne permite o apreciere conformă cu realitatea, ferindu-ne de o „inducție completă", care în statistică ne conduce de multe ori la indicarea unor soluții practice nerealiste.

Din cele prezentate pînă acum, s-a format, credem, o idee asupra a „ce este statistica și ce vrea ea". Statistica, în sine, nu vrea nimic. Noi, în orice domeniu am lucra trebuie *să vrem* să folosim metodele statistice acolo unde situația o cere, să învățăm să mînuim procedee și tehnici statistice simple sau să colaborăm cu statisticienii atunci cînd problemele pe care le avem necesită un aparat statistic mai complex.

Să nu uităm însă niciodată că statistica, în ultimă instanță, ca orice altă ramură a cunoașterii, trebuie privită ca un instrument (foarte interesant, foarte util, dar totuși instrument) în activitatea de zi cu zi, oră de oră, a producției de bunuri materiale și servicii, producție care este strîns legată de civilizația modernă.

ANEXA

DESPRE STATISTICĂ

Mulți oameni de știință și-au expus, cu diferite ocazii, părerile și concepțiile lor privind locul, rolul și importanța statisticii în ansamblul cunoașterii umane. Este desigur dificil să se redea o imagine completă a acestui aspect — mai ales că statistica este o știință în plină evoluție —, problemele de filozofie a statisticii fiind mereu deschise.

În cele ce urmează, vom retranscrie cîteva din ideile unor cunoscuți oameni de știință cu privire la modul lor de gîndire asupra unor trăsături specifice ale statisticii. Alegerea este desigur subiectivă. Un fapt trebuie însă remarcat: toți acești mari statisticieni au relevat, cu orice prilej, importanța practicii, a utilității sociale a metodelor statistice, desolidarizîndu-se întotdeauna de jocul steril al științei de dragul ei însăși.

1. **Anderson, T. W. și Sclove, S. L.**: Statistica intră aproape în fiecare fază a vieții noastre. Buletinul de știri de la radio poate începe cu timpul probabil și se poate sfîrși cu analiza situației economice. Statistica — cu ajutorul unor mijloace specifice —, furnizează baza pentru cercetare în multe domenii ale cunoașterii ca sociologie, fizică, biologie, inginerie, educație, comerț, medicină și jurisprudență.

   Informația asupra unui anumit subiect este sintetizată sub formă de numere; o analiză a acestor numere este făcută cu scopul obținerii unei înțelegeri mai bune a fenomenului ce ne interesează ...

O definiție a statisticii ar putea fi aceasta: extragerea semnificației, a sensului din numere.

(*Introductory Statistical Analysis*, Houghton-Mifflin Co. Boston, 1974, p. 13.)

2. **Bissell, A. F.**: În fine, unul din punctele mele de interes este acela de a încuraja răspîndirea metodelor statistice printre „nestatisticieni" (în locul unui cuvînt mai bun).

Din nou, părerile personale intră în scenă, întrucît regret prăpastia care deseori este atît de adîncă între statistica teoretică și cea aplicată. Beneficiarii potențiali sînt ușor de speriat cu mistica intervalelor de încredere, a coeficienților de regresie, a componentelor dispersiei — pentru a nu da decît cîteva nume — și consider că este de datoria statisticii aplicate să arunce un pod peste această prăpastie. Meritul va fi în întregime al ei, dacă tehnologul înțelege că statistica este creată în scopul de a rezolva probleme reale ce cuprind un element de incertitudine și nu este o mulțime de idei matematice abstracte sau un mijloc de a zăpăci pe cei neinițiați, cu ceva neinteligibil.

(*My kind of statistics*. În: BIAS, vol. 3, nr. 1, 1976, p. 39).

3. **Ciucu, G. și Craiu, V.**: Inferența statistică este acea ramură a metodei științifice care, cu margini specificate de incertitudine exprimate în termeni probabilistici, face trecerea de la observații la concluzii privind populația.

Inferența statistică are două limitări.

Mai întîi, faptul că informația pe care se bazează decizia este de natură statistică, fiind constituită din observații supuse întîmplării.

În al doilea rînd, faptul că se recunoaște explicit nesiguranța concluziei, cu încercarea de a măsura cît mai obiectiv posibil incertitudinea implicată.

În zilele noastre, statistica matematică își diversifică metodele pentru a putea răspunde nenumăratelor întrebări ridicate de cunoașterea realității. Capitole de statistică matematică au căpătat un caracter aproape autonom, așa cum ar fi teoria estimației sau verificarea ipotezelor statistice, iar altele s-au dezvoltat sub impulsul practicii industriale.

Inferența statistică fiind perfectibilă și reflectînd nevoia de certitudine a oamenilor de știință, rămîne o metodă actuală și pentru cei ce se ocupă de logica cercetării științifice.

Contactul statisticii matematice cu alte discipline de interes general a condus la realizarea unor zone de frontieră și la constituirea unui corp de metode care a redus continuu îndoiala omului de știință și i-a astîmpărat setea de certitudine.

(*Inferență statistică*, Editura didactică și pedagogică, București, 1974, p. 1).

4. **Craiu, V.**: Cu fiecare zi viața este influențată din ce în ce mai mult de deciziile bazate pe informații cantitative.

... Metodele statistice moderne, bazate pe teoria probabilităților, sînt de neînlocuit în fizică și biologie, în economie și sociologie, în medicină și agricultură, în psihologie și educație, în industrie și în administrația de stat.

Astronomul stabilește următoarea poziție a unei comete pe baza metodelor statistice; asigurările își bazează activitatea pe tabelele de mortalitate și pe înregistrările statistice; cercetătorii din agricultură și medicină stabilesc dacă o nouă metodă, un nou tratament este superior, pe baza considerațiilor statistice; dezvoltarea sistemului energetic nu poate avea loc fără cunoașterea datelor statistice asupra cererii de energie electrică; inginerii găsesc de neînlocuit teoria selecției în controlul calității producției; administrația de stat își bazează acțiunile pe date statistice.

Metodele statistice ... pot fi considerate unelte care în mîinile unor oameni pricepuți și aplicate la situații pentru care au fost create, dau rezultate folositoare.

(*Verificarea ipotezelor statistice*, Editura didactică și pedagogică, București, 1972, p. 4).

5. **Cox, D. R.**: Statistica se poate aborda din două puncte de vedere foarte diferite. Un matematician poate întreba: are acest domeniu probleme matematice captivante și dă naștere la idei matematice interesante prin ele însele? Statistica teoretică a devenit mult mai matematizată în anii din urmă și cred că toți cei care lucrează în acest domeniu găsesc o largă varietate de ramuri ale matematicii implicate în statistică; fără îndoială, nu există nici o criză de probleme dificile.

Putem însă aborda statistica din punctul de vedere al aplicațiilor. Probleme, care într-un anume sens sînt „statistice" apar virtualmente în toate ramurile științei și tehnologiei, ca și în alte domenii ca, de exemplu, în comerț și administrația de stat.

Aceste probleme sînt deseori foarte dificile și se poate întreba: de fapt, cît de utilă este teoria statistică, în special părțile ei mai recente, în rezolvarea concretă a acestor probleme?

Această a doua întrebare este foarte importantă și răspunsul este greu de dat. De multe ori, nu este de loc ușor să se spună precis care este contribuția unei teorii la rezolvarea unei probleme specifice.

Progresul depinde încă de interacțiunea constantă dintre teorie și practică.

(*Some statistical concepts and their applications*, Imperial College of Science and Technology, Univ. of London, Inaugural lecture, 9 may, 1967, p. 3).

6. **Deming, W. E.**: Poziția competitivă a multor produse japoneze, în conformitate cu însăși mărturia producătorilor lor, a fost cucerită, în mare parte, datorită înțelegerii și folosirii controlului statistic într-un sens foarte larg.

Tehnicile statistice nu au fost în totalitate cauza a ceea ce s-a întîmplat ... dar ele au jucat, în mod sigur, un rol important în acest miracol ...

(*What happened in Japan?* În: „Industrial Quality Control", vol. 24, nr. 2, 1967, p. 90).

7. **Deming, W. E.**: Japonezii nu au cunoscut niciodată controlul statistic al calității decît în sensul larg introdus de Shewhart.

El a fost definit în limba obișnuită cam prin anul 1950 și de atunci încoace scris cu litere mari în felul următor:

CONTROLUL STATISTIC AL CALITĂȚII ESTE APLICAREA PRINCIPIILOR ȘI METODELOR STATISTICE ÎN TOATE FAZELE PRODUCȚIEI, CU SCOPUL DE A PRODUCE ÎN MOD ECONOMIC UN PRODUS DE MAXIMĂ UTILITATE ȘI CARE ARE O PIAȚĂ DE DESFACERE.

(idem, op. cit., p. 91).

8. **Deming, E. W.**: În viața reală nu există constanță veșnică. Există totuși un lucru constant și anume un sistem constant de cauze. Rezultatele produse de un sistem constant de cauze, variază, și, de fapt, ele pot varia între limite largi sau înguste.

Ele variază, dar pun în evidență o caracteristică importantă numită stabilitate. De ce să aplicăm termenii de constanță și stabilitate unui sistem de cauze ce generează rezultatele care variază? Deoarece, o aceeași proporție a acestor rezultate ce variază, continuă să cadă între două limite bine stabilite, oră după oră, zi după zi, atîta timp cît sistemul constant de cauze continuă să opereze. Deci repartiția rezultatelor este constantă sau stabilă.

Cînd un proces de producție se comportă ca un sistem constant de cauze, iar rezultatele lui pun în evidență stabilitatea, se spune că acest proces este într-o stare de control statistic.

(*Some principles of the Shewhart methods of quality control*. În: „Mech. Eng.," vol. 66, 1944, p. 173)

9. **Deming, W. E.**: Calitatea poate fi activitatea fiecăruia într-o întreprindere, dar nimeni nu poate lua parte efectivă într-un program pentru calitate, fără reguli de acțiune bazate pe aplicarea metodelor statistice. Lucrătorii de toate categoriile și de toate funcțiile trebuie să știe cînd acțiunea va fi eficientă și mai ales ce fel de acțiune trebuie să întreprindă.

(*Some statistical logic in the management of quality*. În: „Proc. All India Conference on Quality Control", march, 1971, p. 98).

10. **Deming, W. E.**: În aceste cărți și articole el (Shewhart — n.n.) a explicat foarte clar faptul că producția de masă depinde de standardul de calitate ce convine cerințelor consumatorilor.

Una din marile lui contribuții a fost aceea de a arăta că dimensiunile exacte sînt un mit; nu există astfel de lucruri.

După cum arată Shewhart, oamenii au încercat să construiască dimensiuni prin care un diametru, de exemplu, să se găsească între 2,000 cm și 2,002 cm. Aceasta a însemnat un progres, dar a fost foarte greu de aplicat și nu a condus la o producție economică.

Următorul pas l-a constituit viziunea lui Shewhart: producția economică depinde de atingerea unei repartiții stabile în cadrul specificațiilor. Lucrările lui Shewhart descriu instrumentele statistice care ajută la atingerea unei repartiții stabile (controlul statistic).

. . . .

Au urmat apoi unele noi concluzii pe care le-am predat în Japonia, în 1950 și mai tîrziu, care au ajutat industria de acolo.

De exemplu, a devenit clar că:

1) toate variațiile sînt costisitoare și trebuie reduse,
2) metodele statistice separă cauzele variațiilor in două părți principale:

cauze care apar din sistem și care trebuie corectate de către personalul de conducere,	cauze speciale (care pot fi corectate de către muncitor sau de către un grup de muncitori pe baza unor semnale statistice).

Cauzele speciale au fost numite de Shewhart cauze sistematice. (Eu prefer termenul de cauze speciale deoarece sînt specifice operatorului sau unui grup relativ mic de operatori.)

Operatorul își desfășoară activitatea intr-un anumit mediu. Acest mediu este coordonat de către conducere. Progresul către o producție economică, ca și progresul in controlul statistic al calității depinde în mare măsură de acțiunile pe care le ia personalul de conducere pentru găsirea cauzelor de variație și a cauzelor proastei calități care sînt datorate sistemului, și pentru corectarea sistemului.

(Scrisoare din 29 martie 1976 către V. Gh. Vodă și D. V. Iliescu).

11. **Deming, W. E.**: Anumite procese din natură evidențiază uneori o stare de control statistic. Dezintegrarea radioactivă este un astfel de exemplu. Repartiția timpilor de funcționare fără defecțiuni a tuburilor cu vid etc. furnizează alte exemple.

Dar o stare de control statistic nu este o stare naturală a proceselor industriale. Ea este o cucerire la care se ajunge prin

eliminarea, una cîte una, printr-un efort susținut, a cauzelor speciale de variație succesivă.

(*On some statistical aids toward economic production*. În: „Interfaces", vol. 15, nr. 4, 1975, p. 1).

12. **Deming, W. E.**: După cum spunea Shewhart, cerințele de cunoaștere și măiestrie in industrie și în activitățile publice sînt mult mai severe decît cerințele din știința pură. El ar fi trebuit să adauge că cerințele pentru practica statistică sînt pe departe mult mai mari decît cele necesare predării statisticii.

(*On probability as a basis for action*. În: „The American Statistician", vol. 29, nr. 4, p. 146).

13. **Deming, W. E.**: Proiectarea unui produs este una din aplicațiile importante ale controlului statistic al calității. Nici un efort din partea muncitorilor și nici toată statistica, aplicată pentru detectarea cauzelor speciale de variație cu scopul de a-l ajuta pe muncitor, nu va putea suplini erorile din proiectare.

Greșelile în proiectare sînt greșeli de coordonare (s.n.)

(recenzia lucrării *Statistical Quality Control*, 4th Ed., de E. L. Grant și R. S. Leavenworth, Mc. Graw-Hill Book Co., New York, 1972, publicată in „J. Amer. Statist. Assoc." vol. 68, nr. 343, 1974, p. 748).

14. **Sir R. A. Fisher**: În decursul prezentului secol, pare să se fi făcut un progres foarte mare în domeniul interpretării datelor de observație pentru obținerea unei înțelegeri mai bune a lumii înconjurătoare.

Cele trei aspecte de importanță principală pentru acest progres au fost — primul — folosirea unui aparat matematic mai evoluat și a unor idei mai profunde în statistica matematică; aceasta a condus la metode mai exacte de calcul aplicate la setul de date considerat, care cuprinde toată informația numerică disponibilă asupra subiectului luat în studiu.

Al doilea, întrucit metodele de sintetizare și deducere a concluziilor corecte au devenit adecvate, a fost posibilă apariția unei noi discipline — planificarea experimentelor — care are drept scop obținerea mai completă și mai precisă a datelor și de a evita irosirea unor eforturi in acumularea unor observații greșit planificate sau nerelevante.

Al treilea, ca o consecinţă naturală sau chiar inevitabilă a primilor doi factori, a fost atingerea unei mai simple şi mai complete înţelegeri a structurii şi particularităţilor logicii inductive — adică a raţionamentului: de la selecţie la populaţia din care această selecţie a fost extrasă, de la consecinţe la cauze sau în termeni de logică de la particular la general.

(*Statistical methods and scientific inference.*, În: „J. Roy. Statist. Soc.", Ser. B., vol. 17, nr. 1, 1955, p. 69—70).

15. **Sir R. A. Fisher**: Acum, controlul de recepţie al produselor joacă un rol important în lumea modernă. Cînd o mare întreprindere cum este Royal Navy primeşte materiale de la o firmă producătoare, acestea sînt supuse unei inspecţii şi unei testări atente pentru a reduce frecvenţa acceptării unor produse defecte.

Instrucţiunile pentru cei ce efectuează acele testări sînt astfel alcătuite încît să menţină la un nivel cît mai scăzut costul testărilor şi frecvenţa respingerii loturilor satisfăcătoare.

Este necesară multă ingeniozitate şi mult talent pentru a face aceste procedee de control de recepţie într-adevăr efective şi economice.

(idem, op. cit., p. 70).

16. **Good, I. J.**: Matematicianul, statisticianul şi filozoful folosesc în diferite moduri o teorie a probabilităţilor.

Matematicianul dezvoltă consecinţele ei formale, statisticianul aplică rezultatele matematicianului, iar filozoful descrie în termeni generali în ce constă această aplicare.

Matematicianul creează instrumente simbolice fără a se îngriji prea mult de utilitatea lor; statisticianul foloseşte aceste instrumente; filozoful vorbeşte despre ele. Fiecare îşi face munca mai bine dacă ştie cîte ceva despre munca celorlalţi doi.

(*Kinds of probability*. În: „Readings in Applied Statistics", edited by W. S. Peters, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1969, p. 29).

17. **Gupta, B. N.**: Metodele statistice sînt inductive prin natura lor, deoarece generalizările rezultă din observaţii individuale.

Generalizările deduse dintr-o investigaţie statistică arată că aceste generalizări sînt adevărate „în medie". Ele pun în

evidenţă numai comportamentul tipic al tuturor obiectelor luate în studiu, dar nu descriu comportamentul elementelor luate separat, deoarece există o stabilitate mai mare în colectivitate decît în individ.

(*An Introduction to Modern Statistics*, Bookland Private Ltd., Calcutta, 1962, p. 10).

18. **Hotelling, H.**: Un bun profesor de statistică trebuie să fi urmat un curs de baze matematice, care să fi inclus cel puţin cîteva noţiuni de teoria funcţiilor şi de geometrie euclidiană n-dimensională. De asemenea, algebra şi analiza matematică sînt utile, ca şi cîteva elemente de geometrie diferenţială.

Dar, oricîtă matematică de acest fel s-ar acumula, ea nu constituie prin ea însăşi o condiţie suficientă pentru a oferi calitatea de profesor de statistică.

Lucrul cel mai important este acela de a cunoaşte însăşi teoria statistică în adîncime, de la partea cea mai elementară pînă la cea mai elevată, incluzînd partea matematică a metodologiei şi o înţelegere clară a modului în care aceasta se aplică în diferite domenii specifice.

Alături de matematica pură şi de cunoaşterea teoriei statistice, un statistician competent sau un profesor de statistică are nevoie de o legătură foarte intimă cu problemele dintr-unul sau mai multe domenii practice. Acest lucru este foarte important.

A fost nevoie ca excelenţi matematicieni ... să inducă în eroare studenţii pînă la eşec, pentru a căpăta această înţelegere, pentru aplicaţii, care este necesară pentru o activitate statistică adecvată.

(*The teaching of statistics, Address at the meeting of the Institute of Mathematical Statistics at Hanover, N. H., september 10, 1940.* În: „Ann. Math. Statist.", vol. 8, 1940, p. 463).

19. **Hammersley, J. M.**: ... Descoperirile şi invenţiile sale (ale lui Sir R. A. Fisher — n.n.) în domeniul tehnicilor statistice au izvorît din probleme practice şi pămînteşti foarte complicate.

Multe din acestea erau pămînteşti chiar în sensul literal al cuvîntului, fiind create ad-hoc pentru experimentele de agricultură desfăşurate la Rothamsted. Altele au apărut din interesul său pentru genetică, biologie, astronomie, meteorologie, geofizică etc.

El a menținut o interacțiune constantă între statistica teoretică și statistica aplicată, hrănind mereu pe una cu cealaltă.

. . . .

Fisher este propriul său contraexemplu — și cel mai bun — la afirmația că tehnicile statistice au izvorît din apele liniștite ale teoriei.

(*Statistical tools*. În: „The Statistician", vol. XXII, nr. 2, 1974, p. 90).

20. **Ionescu, H. M.**: Sarcina statisticii matematice constă în elaborarea metodelor care să permită descoperirea trăsăturilor fundamentale ale fenomenului sau procesului de masă cercetat. În acest scop, se folosește un material, de obicei restrîns, obținut pe cale experimentală, adică prin efectuarea unor probe, în general independente între ele. Desigur că acest material prezintă, în primul rînd, un interes practic imediat, deoarece el ne furnizează informații asupra fenomenului sau procesului observat; așadar, pe baza lui putem caracteriza un anumit proces concret.

. . . .

Problemele statisticii matematice apar însă atunci cînd pe baza materialului obținut prin experiență trebuie să caracterizăm fenomenul în ansamblul său...

(*Statistica matematică*, Editura de Stat Didactică și Pedagogică, București, 1962, p. 11).

21. **Iosifescu, M. ș.a.**: Dezvoltarea vertiginoasă a teoriei probabilităților corespunde creșterii importanței și interesului față de studiul legilor statistice în toate domeniile științifice, mai ales cînd este vorba de structura lor logico-inductivă. Explicarea unor fenomene cu ajutorul acestor legi — nicidecum opuse legilor deterministe — a implicat constituirea unui nou limbaj matematic, acela al probabilității.

(*Teoria probabilităților și statistica matematică*, Editura tehnică, București, 1966, p. 5).

22. **Iosifescu, M. ș.a.**: Nu există științe care să nu aibă legi cu caracter statistic. Utilitatea sporește însă pe măsura dezvoltărilor lăuntrice teoriei, pe măsura apariției de noi concepte generalizatoare...

(idem, op. cit., p. 6).

23. **Kanji, G. M.**: Cea mai mare dificultate pentru cei care studiază statistica este transferul de gîndire de la cunoștințele lor de bază în matematică, la ideile de bază ale statisticii.

Este greu pentru ei să înțeleagă că, într-un anume sens, acolo unde matematica se termină, începe statistica.

(*The role of statistical laboratory in the teaching of statistics*. În: „International Journal of Mathematical Education in Science and Technology", vol. 5, nr. 1, 1974, p. 54).

24. **Lord Kelvin**: Atunci cînd se poate măsura ceea ce ne interesează și exprima în numere, atunci cunoaștem ceva despre fenomenul luat în considerare, dar cînd nu putem măsura aceasta și nici n-o putem exprima în numere, cunoașterea noastră este palidă și nesatisfăcătoare.

(citat după Tippett, L. C., H. *Statistics*, 3d edition, Oxford Univ. Press, 1968, pp. 2–3).

25. **Kolmogorov, A. N.**: Întrucît rolul important al legilor statistice este incontestabil, se pune problema metodelor lor de studiu.

Prima care se pune este ideea posibilității de a le cerceta pur empiric, experimental. Deoarece o lege de probabilitate se manifestă în procese de masă, este natural ca, pentru descoperirea ei, să fie necesară efectuarea unui număr mare de experiențe.

Această concepție este numai în parte adevărată. După ce am stabilit experimental unele legi de probabilitate, putem deduce din ele noi legi de probabilitate, pe cale logică sau prin calcul, cu ajutorul unor ipoteze generale.

(*Matematica, conținutul, metodele și importanța ei*, Editura științifică, București, 1960, vol. II, cap. XI, p. 324).

26. **Kolmogorov, A. N.**: ... întreaga teorie matematică a probabilităților poate fi construită din punct de vedere formal ca o teorie a măsurii, cu singura ipoteză specială că măsura „întregului spațiu" U este egală cu unitatea.

Cu toate acestea, teoria probabilităților rămîne o disciplină matematică de sine stătătoare prin esența problemelor

tratate; rezultatele fundamentale ale teoriei probabilităţilor par artificiale şi inutile din punctul de vedere al teoriei pure a măsurii.

(idem, op. cit., p. 349).

27. **Mihoc, Gh.**: Activitatea economică din zilele noastre, desfăşurată într-o tot mai strînsă interdependenţă între ţări şi popoare, obligă fiecare naţiune să valorifice la maximum toate resursele de care dispune.

Aplicaţiile matematice şi statistico-matematice care se introduc tot mai mult în diferite ramuri ale economiei naţionale învederează un pas foarte important în conducerea ştiinţifică a producţiei de bunuri şi servicii, contribuind la progresul economic al ţării. Acest progres are un netăgăduit caracter calitativ.

....

Dar metodele matematice aplicate în economie au nevoie, într-o covîrşitoare măsură, de o informaţie cît mai complexă şi mai precisă pentru a fundamenta eficienţa revoluţiei tehnico-ştiinţifice din zilele noastre, bazată pe calculul economic, pe prelucrarea electronică a datelor, pe automatizare, pe programare şi pe cibernetică.

Această informaţie se reflectă în statistica matematică, ale cărei metode şi posibilităţi tehnico-ştiinţifice fundamentează astăzi analiza, previziunea şi conducerea fenomenelor şi circuitelor de producţie economică.

(*Rolul statisticii matematice în realizarea obiectivelor economice*. În: „Metode noi şi probleme de perspectivă ale cercetării ştiinţifice", Editura Academiei R. S. România, Bucureşti, 1970, p. 269).

28. **Mănescu, M.**: O metodă modernă de control tehnic al calităţii care a dat rezultatele cele mai bune şi este intens folosită în ţările cu industrie dezvoltată, în producţia de serie şi de masă, este controlul statistic.

....

Aplicarea controlului statistic permite, de asemenea, realizarea unor economii de materiale, prin micşorarea procentului de rebuturi; întărirea disciplinei controlorilor, care urmăresc permanent starea calităţii producţiei, angrenîndu-se organic în procesul tehnologic; îmbunătăţirea condiţiilor de proiectare a proceselor tehnologice, deoarece datele înregistrate pe fişele de control statistic pe fiecare maşină-unealtă sînt folosite în aprecierea posibilităţilor acesteia de a executa lucrări de o anumită precizie, în cadrul unor anumite toleranţe, permiţînd totodată distribuirea concretă a lucrului pe maşini.

Pe baza fişelor de control se poate stabili timpul optim pentru reparaţii curente şi capitale ale utilajelor, se poate cunoaşte în orice moment felul cum decurge procesul de producţie în întreprindere şi aprecia în mod obiectiv indicii de calitate realizaţi.

(„Introducere" la lucrarea *Analiza statistico-matematică a calităţii producţiei industriale* de N. Rancu şi L. Tövissi, Editura ştiinţifică, Bucureşti, 1964, p. 7–8).

29. **Niculescu, D.**: Folosirea metodelor statisticii matematice şi ale calculului probabilităţilor în controlul tehnic al calităţii permite introducerea gestiunii moderne a calităţii.

Pe lîngă stavila pe care o pune livrării unor produse de calitate scăzută, controlul statistic furnizează fundamentarea ştiinţifică a conducerii active a proceselor de producţie, permiţînd depistarea tendinţelor şi a cauzelor perturbatoare, stabilirea acţiunilor care trebuie întreprinse.

(*Conducerea ştiinţifică a calităţii*. În: „Calitatea Producţiei şi Metrologie", vol. III, nr. 1, 1973, p. 3).

30. **Narula, S. C.**: În general, problemele practice sînt o bună sursă pentru cercetarea fundamentală şi aplicativă în toate domeniile.

Întrucît aceste probleme nu sînt disponibile în mediile academice, este recomandabil ca toţi cercetătorii să dezvolte legături foarte strînse cu lucrătorii din diverse „industrii" pentru a vedea dacă există probleme de interes mutual.

(*Systematic ways to identify research problems in statistics*. În: „International Statistical Review", vol. 42, nr. 2, p. 205).

31. **Onicescu, O. şi Mihoc, Gh.**: Statistica matematică este o ramură a ştiinţelor matematice care se ocupă cu noţiunile şi cu meto-

dele specifice studiului colectivităților. Bazată pe teoria probabilităților, din care s-a desprins, încă de multă vreme, statistica matematică, oferă științei metode generale decurgînd din caracterul abstract al relațiilor cu care se ocupă. În esență, rolul statisticii matematice este de a determina, pe baza datelor empirice, informații cît mai precise asupra repartiției probabilistice a fenomenului de masă care ne interesează.

(*Statistica matematică — instrument al științelor naturii și societății*. În: „Cercetările multidisciplinare și interdisciplinare, originea, dezvoltarea și perspectivele lor", Editura Academiei R. S. România, București, 1972, p. 25).

32. **Shewhart, W. A.**: Contribuția în timp a statisticii depinde nu atît de mult de aportul în industrie al unui număr mare de statisticieni de înaltă clasă, cît de crearea unei gîndiri statistice la fizicieni, chimiști, ingineri și la alții care, într-un fel sau altul, vor avea în mînă dezvoltarea și dirijarea proceselor de producție ale viitorului.

(*Statistical Method from the Viewpoint of Quality Control*, Edited by W. E. Deming, The Graduate School, Department of Agriculture, Washington, D.C., 1939, p. 13).

33. **Zelen, M.**: Ce este știința statistică? Statistica, teoria probabilităților și analiza datelor de observație sînt importante într-o varietate largă de activități științifice și tehnice, ca și în viața de toate zilele.

Teoria probabilităților furnizează limbajul general pentru descrierea și construirea teoriilor despre incertitudine, iar statistica — tehnicile pentru adecvarea modelelor probabilistice datelor experimentale și pentru luarea deciziilor în condiții de incertitudine. Calculul statistic, la rîndul lui, furnizează algoritmii și codurile de efectuare a calculelor, pentru a răspunde în mod corect și eficient la sarcinile puse în fața cercetătorului, de analiza datelor experimentale. Știința statisticii este interacțiunea dintre statistică, știința calculatoarelor și importante aplicații practice.

(*Statistical Science*, Division Guidebook, State University of New-York at Buffalo, 1974).

34. **Wilks, S. S.**: Ceea ce avem nevoie în statistică sînt cursuri elementare, la nivel elementar, în care studentul să se poată concentra asupra conceptelor de bază, prin efectuarea unui număr mare de probleme și exerciții de laborator pentru fixarea acestor idei fără să se piardă în manipularea fără sens a unor formule.

(*Undergraduate statistical education*. În: „J. Amer. Statist, Assoc"., vol. 46, 1951, p. 1).

35. **Wilks, S. S.**: Populațiile statistice, în sensul modern, sînt, în general, generate de operații cu caracter repetitiv, fie pe cale naturală, fie efectuate de om. Speciile de plante și animale sînt populații naturale. Revoluția industrială cu tehnicile ei de producție de masă, procese automate, piese interschimbabile și probleme de marketing, a creat o varietate fără sfîrșit de universuri statistice, fiecare dintre ele capabil de a fi investigat cu ajutorul metodologiei statistice.

Ipotezele care stau la baza metodelor statistice moderne sînt, probabil, mult mai complet satisfăcute din punct de vedere probabilistic de populațiile generate în știință și industrie decît cele pe care le întîlnim în domeniul social și economic.

(*The rise of modern statistical science*. În: „Proc. MIT Industrial Statistics Conference", Pitman Publ. Co., New York, 1939, p. 283).

BIBLIOGRAFIE SELECTIVĂ

Ciucu, G., Craiu, V., *Teoria estimației și verificarea ipotezelor statistice*, Editura didactică și pedagogică, București, 1968.

Ionescu, C., *Perfecționarea conducerii și planificării economiei naționale și sarcinile statisticii*. În: „Revista de statistică", nr. 10, 1967, pp. 3–6.

Mihoc, Gh., Urseanu, V., *Legea numerelor mari — element component al legității statistice*. În: „Revista de Filozofie", tom. 12, nr. 6, pp. 24–36.

Mihoc, Gh., Craiu, V., *Teoria probabilităților și statistica matematică*, Editura didactică și pedagogică, București, 1970.

Rebedeu, I., *Politica Partidului Comunist Român în domeniul învățămîntului, științei și culturii*. În: „Era socialistă", nr. 3, 1975, pp. 31–37.

ÎN LIMBI STRĂINE:

Dunin-Barkovski, I. V., Smirnov, N. V., *Teoria veroiatnostei i matematiceskaia statistika v tehnike*, G.T.T.I., Moskva, 1955.

Firkowicz, Sz., *Statystyczne Badanie Wyrobow*, W.N.T., Warszawa, 1970.

Sachs, L. *Statistische Auswertungsmethoden*, Springer Verlag, Berlin, 1972.

CUPRINS